Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Informática

Dissertação de Mestrado

Mestrado em Engenharia Informática

Extracção e Representação de Metadados num Contexto de Produção de Vídeo

João David Godinho Moreira Mateus (27231)

Lisboa
(2010)

Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia
Departamento de Informática

Dissertação de Mestrado

Extracção e Representação de Metadados num Contexto de Produção de Vídeo

João David Godinho Moreira Mateus (27231)

Orientador: Prof. Doutor Nuno Manuel Robalo Correia

Trabalho apresentado no âmbito do Mestrado em Engenharia Informática, como requisito parcial para obtenção do grau de Mestre em Engenharia Informática.

Lisboa
(2010)

Aos meus pais.

"..let's say knowledge is a tree
it's growing up just like me.."

Agradecimentos

Gostaria de começar por agradecer ao meu orientador, Prof. Nuno Correia, por toda ajuda na elaboração desta dissertação. A sua orientação, através da sua disponibilidade para a troca de ideias, espírito crítico e dedicação, foram essenciais para guiar este trabalho.

Queria agradecer à Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT/UNL) que através do Centro de Informática e Tecnologias de Informação (CITI) disponibilizou as condições para a realização deste trabalho e à Agência de Inovação pelo apoio financeiro no âmbito do projecto VideoFlow. Ao parceiro Duvideo, por apresentar um trabalho científico interessante e aliciante, assim como ao Rui Jesus pela cooperação na integração de tecnologias.

Por fim, gostava de agradecer ao meus colegas e amigos António Costa, Luís Silva, Ricardo Mateus, Rui Chambel, Tiago Ribeiro, Diogo Rendeiro, José Sousa e Ricardo Perleques por todas as discussões que nem sempre levaram a uma conclusão mas que sempre trouxeram boa disposição. Aos meus pais e à Nélia, por toda a confiança, apoio e ajuda, sem os quais tudo teria sido muito mais difícil.

Resumo

Cada vez mais é necessário aproveitar e reutilizar os conteúdos dos arquivos de vídeo de forma a extrair informação que possa tornar os processos de trabalho mais eficientes. No contexto de produção de conteúdos vídeo, os arquivos necessitam de técnicas de extracção de metadados, incluindo detecção automática de cenas, identificação de pessoas, objectos e conceitos. Este projecto tem o objectivo de desenvolver e testar técnicas para resolver estas necessidades, de modo a enriquecer a semântica dos conteúdos e que possam ser utilizados para futuras pesquisas. Estes metadados serão incorporados em formatos normalizados, nomeadamente o Material eXchange Format (MXF) - visto ser este o formato utilizado pela Duvideo, empresa que participa no projecto VideoFlow onde esta tese está integrada. Além da extracção de metadados, são também propostas interfaces de modo a que o acesso e visualização dos conteúdos, seja feita de uma maneira mais eficaz pelo utilizador final.

Palavras-chave: Produção de Vídeo, Metadados, Segmentação de Vídeo, Identificação de Objectos, Detecção de Conceitos

Abstract

The reuse of video content from archives, is a task that needs to receive more attention with the goal to extract information that can make workflows more efficient. In the context of video production, there is a need for techniques for automatic scene detection, identification of persons, objects and concepts. This project aims to develop and test techniques, in order to increase the semantic content in a way that it can be used for search engines. This metadata will be incorporated into standard formats. We will give more focus to Material eXchange Format (MXF), since this is the format used by Duvideo, a company that participates in the project VideoFlow, the project that sets the scope for this thesis. Regarding the extraction of metadata, interfaces are also proposed so that access and visualization are made in a more effective way to the end user.

Keywords: Video Production, Metadata, Video Segmentation, Object Identification, Concepts Detection

Conteúdo

Lista de Figuras

Lista de Tabelas

Listagens

1

Introdução

Com a democratização do acesso de alta velocidade à Internet, os conteúdos disponíveis, que anteriormente eram essencialmente textuais, começam agora a ter uma componente multimédia muito mais forte. O YouTube é um marco nesta área sendo possível observar o seu grande crescimento tanto a nível de utilizadores como de conteúdos. Um utilizador normal pode criar o seu canal televisivo e mostrar a informação que desejar. Os canais televisivos começam também a disponibilizar grande parte dos conteúdos nos seus *sites*, mas também em *sites* sociais como o YouTube. Estes conteúdos podem ser um complemento ao que é transmitido pela televisão ou então uma replicação total ou parcial do que é emitido.

Com estas novas necessidades, novos problemas acabam por surgir e quando olhamos para esta mudança de perspectiva na forma como o conteúdo é disponibilizado, os modos de pesquisar, aceder e visualizar necessitam de ser repensados.

Este documento foca-se nas necessidades de produção de conteúdos vídeo, de forma a criar soluções que permitam facilitar e reutilizar o trabalho anterior.

1.1 Motivação

Com o aumento da quantidade de horas de vídeo que fazem parte dos arquivos das empresas de produção de vídeo, é necessário encontrar maneiras de reutilizar estes conteúdos. Cada vez mais se espera que o produto final chegue ao consumidor o mais rapidamente possível, o que faz com que exista cada vez menos tempo para criá-lo.

A aproximação para realizar este projecto será através de uma abordagem a métodos de análise de imagens de modo a enriquecer a informação extraída dos vídeos, tanto em qualidade como em quantidade. Com esses novos elementos (metadados) encontrados, podemos identificar novas formas de pesquisa e navegação através de interfaces que tirem proveito de toda a indexação realizada. É dado um melhor uso aos arquivos de vídeo existentes e todo o processo de produção de conteúdos se torna mais rápido e eficiente.

1.2 Descrição do problema

Este trabalho encontra-se integrado no âmbito do projecto VideoFlow, em parceria com a empresa Duvideo, financiado pelo QREN (Quadro de Referência Estratégico Nacional), com o objectivo de melhorar *workflows* de produção de conteúdos audiovisuais.

O *workflow* tradicional de produção de conteúdos da Duvideo, pode ser observado na figura 1.1.

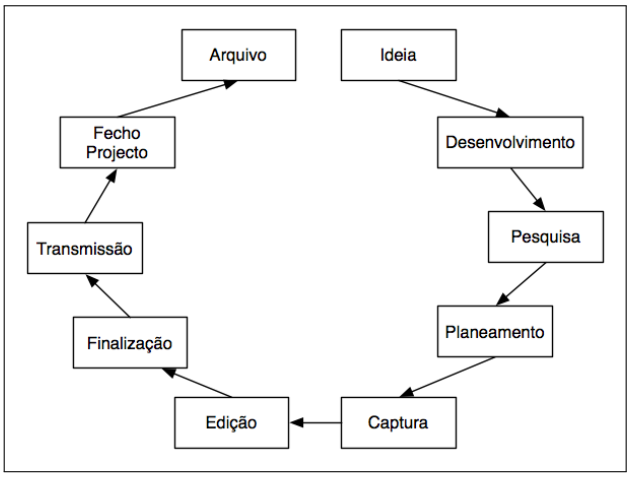


Figura 1.1: Ciclo normal da produção de conteúdos audiovisuais

Este *workflow* inclui passos que requerem recursos limitados e com custos associados à sua utilização. É o caso da Captura, onde é necessário enviar uma equipa e material para o local onde se vai filmar. Devido ao Arquivo não estar preparado para responder a formas de pesquisas mais complexas, este não é reutilizado no máximo das suas capacidades. Empresas que tenham um arquivo de vídeo de dimensões consideráveis, sentem a necessidade de arranjar métodos de forma a que a Pesquisa possa tirar partido do Arquivo, de modo a que processos como a Captura, não necessitem de realizar trabalho que já possa ter sido feito. Deste modo, procura-se uma reutilização eficaz do material existente.

A procura de fragmentos de vídeo relevantes numa colecção é uma tarefa mais complicada do que a de pesquisa num domínio textual. O objecto de procura é mais facilmente representado numa linguagem natural do que através de conceitos de alto nível, no entanto isto requer criar extensas descrições dos conteúdos. O tipo mais normal de informação que é adicionado aos vídeos são os descritores textuais. Estes caracterizam o vídeo na totalidade ou secções do mesmo. Estes descritores são inseridos pelo observador após uma visualização do vídeo em questão, o que torna todo processo bastante moroso, assim como susceptível de uma interpretação subjectiva do conteúdo apresentado.

"One Picture is Worth Ten Thousand Words" (Fred R. Barnard)

Estes métodos devem ser realizados duma forma o mais automática possível, com recurso à computação o que tornará o processo mais célere e com o mínimo de recurso à subjectividade humana. Apesar disto, a validação humana destes resultados será sempre uma parte importante, visto estes não serem perfeitos.

1.3 Solução apresentada

Como forma de tornar este processo de captura de metadados mais célere, através de métodos que tragam resultados o mais precisos possíveis, serão utilizadas técnicas de segmentação para reduzir o espaço de procura e ferramentas de análise de imagem para criar informação a ser usada em futuras pesquisas. Todos os resultados serão integrados com o *workflow* da Duvideo, no formato normalizado Material eXchange Format (MXF) [Dev02]. Esta solução não pretende substituir as anotações manuais mas sim complementá-las de modo a gerar metadados mais ricos semanticamente. Além da componente relacionada com a geração de metadados, esta solução também inclui uma interface cliente para a visualização, validação e utilização da informação gerada.

1.4 Principais contribuições previstas

Com a solução apresentada, prevê-se que a criação de metadados seja realizada de uma forma mais eficiente, com recurso a uma biblioteca de algoritmos, e que permita adicionar mais informação semântica.

Através da disponibilização de uma interface simples de utilizar, será possível proceder a uma avaliação da informação criada e ainda realizar pesquisas que retirem todo o potencial dos metadados gerados.

Deste modo, os utilizadores finais (jornalistas, guionistas, produtores e realizadores) têm um acesso mais rápido aos recursos necessários para efectuar o seu trabalho.

1.5 Organização do documento

Além do presente capítulo, a estrutura deste documento inclui mais três capítulos:

Capítulo 2 - Trabalho relacionado

Este capítulo inicia-se com uma descrição das características dos sistemas de arquivos vídeo (secção 2.1). Serão ainda apresentados os seguintes tópicos relevantes para o projecto:

- Segmentação de vídeo (secção 2.2)
- Análise de imagem (secção 2.3)
- Integração de metadados (secção 2.4)

Capítulo 3 - Solução

Neste capítulo é descrita a solução proposta. Esta descrição é composta pelas seguintes secções:

- Desenho (secção 3.1)
- Realização (secção 3.2)
- Avaliação (secção 3.3)

Capítulo 4 - Conclusões e Trabalho Futuro

Neste capítulo descrevem-se as conclusões referentes à solução proposta e os próximos passos para o trabalho futuro.

- Conclusões (secção 4.1)
- Trabalho futuro (secção 4.2)

2

Trabalho relacionado

Neste capítulo será descrito o trabalho relacionado, que servirá como base para a realização do sistema proposto nesta dissertação. O capítulo inicia-se com uma apresentação de soluções já implementadas para problemas semelhantes em arquivos de vídeo (secção 2.1). A secção seguinte refere-se às técnicas que servirão de suporte para a solução final, nomeadamente ao nível da segmentação de vídeo (secção 2.2) e análise de imagem (secção 2.3). Por fim, são apresentados os formatos mais utilizados para o armazenamento de conteúdos audiovisuais e metadados associados (secção 2.4).

2.1 Sistemas de arquivo de vídeo

Nos arquivos de vídeo tradicionais, o vídeo encontra-se armazenado em cassetes de fita enquanto os metadados já se encontram na maior parte dos casos num suporte digital. Estando em formatos diferentes torna-se difícil a criação de ferramentas para tirar o máximo proveito do material. O passo natural é a conversão para o digital do material mais antigo, juntando-o ao material mais recente que já é criado, na maior parte dos casos, numa base digital.

As estações de televisão e as empresas de produção de conteúdos audiovisuais já iniciaram este processo de digitalização do seu material audiovisual de modo a tirarem maior proveito do mesmo. Este também é o caso da Duvideo, que ainda tem cerca de 90% do seu arquivo num suporte analógico e que pretende iniciar esta conversão. No entanto, este é um processo lento e dispendioso. Com a digitalização destes arquivos,

é fácil encontrar vantagens:

- Cópia sem perda de qualidade.
- Facilidade de acesso a material que se pode encontrar em locais geograficamente distantes.
- Acesso ao material por várias pessoas em simultâneo para a produção de conteúdos.
- Criação de ferramentas de pesquisa, acesso e visualização dos conteúdos tirando partido da melhor ligação entre os metadados e os vídeos.

O foco deste documento será o último ponto, onde as técnicas descritas nas secções seguintes têm um papel relevante.

Estes sistemas de arquivo de vídeo devem oferecer quatro funcionalidades principais [HLMS95] aos seus utilizadores:

Pesquisa

Estes arquivos são acedidos de forma a encontrar vídeos. Estes podem encontrar-se num formato bruto (os vários *takes* de uma reportagem ou de uma determinada cena) ou já como produto final. Normalmente quer-se pesquisar por vídeos de uma determinada pessoa - e.g., vídeos do primeiro-ministro - ou vídeos relacionados com um determinado assunto - e.g., vídeos das eleições legislativas. É essencial fornecer métodos para responder a estas necessidades.

Navegação estruturada

Um dos problemas nos vídeos em suporte analógico é o seu acesso sequencial. Não existem pontos de acesso directos, pelo que se perde muito tempo neste processo. Já com suporte digital existe a possibilidade de um acesso directo a pontos do vídeo, sendo apenas necessário construir uma base de informação de forma a descrever o seu conteúdo.

Navegação por contexto

Por vezes só as imagens não explicam tudo, por exemplo vídeos sobre eleições legislativas e vídeos sobre eleições presidenciais, são semelhantes no conteúdo mas diferentes no seu contexto. É necessário adicionar o máximo de informação (metadados) possível de modo a enriquecer a semântica do vídeo.

Anotações

São as anotações que permitem enriquecer os vídeos, como foi apresentado anteriormente. No entanto estas devem ser mais do que descrições gerais, sendo conveniente que estejam associadas e descrevam segmentos do vídeo, de forma a ajudarem ao acesso directo à cena.

As próximas secções apresentam sistemas de arquivo de vídeo e formas utilizadas para os tentar rentabilizar através de uma melhor relação entre conteúdos e metadados.

2.1.1 Video Storage and Retrieval (VideoSTAR)

O VideoSTAR é o sistema descrito em [HLMS95], que tem como objectivo a partilha de informação para responder às necessidades dos seus utilizadores terem uma visão dos vídeos e dos metadados agregados de forma a facilitar a sua compreensão.

A sua arquitectura pode ser observada na figura 2.1[1]. Esta é composta por:

- Repositórios para os vídeos e os seus metadados
- Um modelo de dados que disponibiliza mecanismos para representar os conteúdos armazenados e relações entre essas entidades.
- Uma API (Application Programming Interface) para carregar os vídeos e aceder a determinadas zonas dos mesmos.
- Ambientes de pesquisa, navegação, acesso e visualização dos conteúdos.

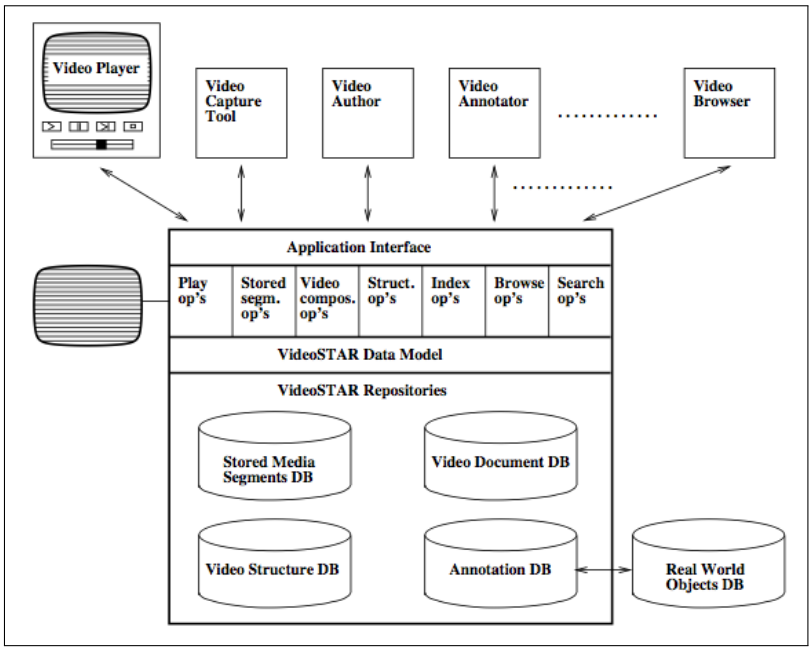


Figura 2.1: Arquitectura do sistema VideoSTAR

[1]Todas as imagens desta secção foram retiradas de [HLMS95].

É disponibilizado ao utilizador um *player* de vídeo que comunica com outras ferramentas, fornecendo informação sobre o vídeo que está a ser visualizado - i.e., a imagem actual. Desta forma, a ferramenta de navegação do documento (Fig. 2.2), que contém os metadados referentes ao vídeo, pode ir actualizando o assunto, indicando onde se encontra no contexto geral do vídeo.

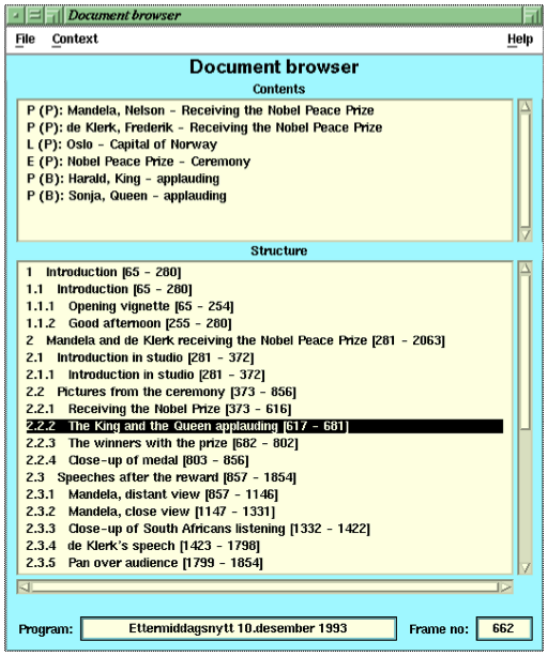


Figura 2.2: Ambiente de navegação.

Neste ambiente de navegação também são apresentados ao utilizador os tipos de conteúdos que se encontram presentes no vídeo. Foram especificados três tipos de categorias para caracterizar os conteúdos: (P) Pessoa, (L) Local e (E) Evento. Assim é possível utilizar estas categorias para realizar pesquisas sobre o arquivo, através do ambiente de pesquisa (Fig. 2.3). Com os resultados obtidos, o utilizador pode aceder directamente ao local do vídeo que está relacionado com a sua pesquisa. O ambiente de anotações de metadados (Fig. 2.4) consome uma parte significativa do tempo dos utilizadores. Visto ser um processo manual, o utilizador deve parar o vídeo sempre que desejar adicionar informação relacionada com um intervalo de tempo no vídeo. O *player* comunica então com o ambiente de anotações a *frame* actual, para dar início ao intervalo. Para fechar o intervalo, o utilizador tem de seleccionar a anotação, que se encontra num modo "aberto", quando o vídeo atingir o ponto onde o contexto se tenha alterado e deixe de fazer sentido para a anotação actual.

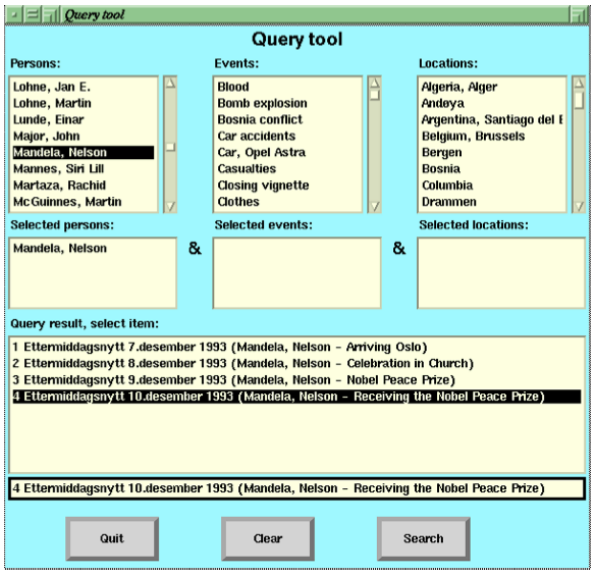


Figura 2.3: Ambiente de pesquisa.

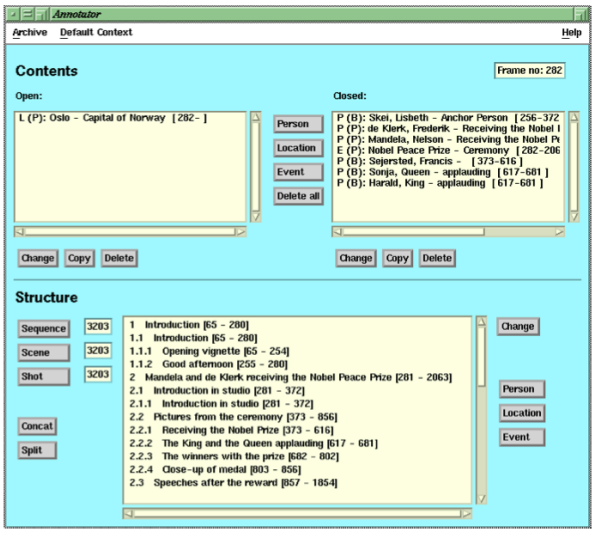


Figura 2.4: Ambiente de anotações.

Nos inquérito realizados aos utilizadores, estes apontaram os seguintes aspectos a melhorar: *a*) aumento na expressividade das pesquisas, adicionando mais categorias; *b*) possibilidade de saber o tamanho dos vídeos relacionados com a pesquisa; e *c*) possibilidade de pesquisa por texto livre.

2.1.2 The Informedia Project (TIP)

O TIP [WCGH99] nasceu na Universidade de Carnegie Mellon, com o objectivo de criar uma biblioteca de vídeo com uma dimensão de um *terabyte* que utilizava técnicas automáticas, ao nível de reconhecimento de voz e de processamento de imagem, para realizar segmentação e indexação dos materiais. A criação de uma interface que permitisse que os utilizadores extraíssem o máximo dos dados gerados também era uma parte importante do projecto.

A biblioteca era composta por dois tipos de vídeos: notícias (1.000 horas) e documentários (400 horas). No final existiam cerca de 40.000 histórias (segmentos) gerados, um número que aumentava diariamente visto a biblioteca também estar em constante crescimento.

A detecção de faces foi conseguida com sucesso, quando a face se encontrava de frente e bem iluminada e assim sendo foi disponibilizada ao utilizador a possibilidade de realizar pesquisas por faces.

Para a detecção de cores a utilização apenas de histogramas não se revelou um bom processo. Foi utilizado um método baseado na percepção humana de agrupar cores [GPF98]. Este agrupamento é realizado primeiro com as cores predominantes e em regiões uniformes. É criado um vector multi-dimensional para cada uma dessas cores e regiões que servirá para criar índices. Para a pesquisa de imagens, o utilizador escolhe a cor e a região na imagem e o sistema devolve as imagens relevantes.

Foi realizado OCR (Optical Character Recognition) sobre as *keyframes* (*frames* que identificam segmentos) para extrair texto presente nas imagens para adicionar informação à descrição do segmento (muito comum nos vídeos de notícias).

Como o processamento automático não é perfeito e introduz ambiguidade, a pesquisa pode introduzir ainda mais ambiguidade caso os utilizadores não saibam como a utilizar para encontrarem o que desejam. O elemento essencial do *design*[2] foi uma disposição alternativa sobre a navegação através de abstracções como cabeçalhos (*headlines*), miniaturas (*thumbnails*), tiras de filmes (*filmstrips*) e resumos (*skims*).

[2]Todas as figuras desta secção foram retiradas de [WCGH99].

Headlines

Na figura 2.5, quando o utilizador passa com o rato sobre o *thumbnail* aparece o *headline* correspondente. Estes contêm as palavras mais importantes[3]. É mostrado ainda o tamanho do segmento e a sua data de criação. Existem no entanto problemas associados aos mesmos - i.e., não são de fácil leitura, visto as palavras não aparecerem naturalmente; não existe uma semântica consistente com o contexto do vídeo.

Figura 2.5: Resultado de uma pesquisa.

Thumbnails

Os *thumbnails* que mostram as *keyframes* dos segmentos, revelaram-se muito úteis para os utilizadores seleccionarem os resultados pretendidos.

[3]É utilizado *term frequency-inverse document frequency* como medida para classificar a importância duma palavra. Isto significa que uma palavra aumenta a sua importância proporcionalmente ao número de vezes que aparece no documento, equilibrando com a frequência de vezes que ela aparece. Deste modo, palavras como "verde" ou "conferência" têm uma maior importância que outras como "de" ou "a".

Filmstrips

As *keyframes* dos segmentos do vídeo podem ser representadas numa ordem sequencial. É possível ver quais são os segmentos onde cada palavra da pesquisa aparece. No exemplo da figura 2.6, observa-se que existe um código de cores para representar cada palavra da pesquisa e cada *thumbnail* é marcado com a cor da palavra que ocorre no próprio. Com a selecção de um *thumbnail*, o vídeo é colocado na posição que é representada pelo *thumbnail*. Existe a opção do utilizador avançar temporalmente no vídeo, saltando de segmento em segmento.

Figura 2.6: Exemplo de um *filmstrip*.

Skims

Os *skims* são como *trailers* dos vídeos, fazendo com que se observe todo o vídeo num espaço de tempo mais reduzido. São mencionados os seguintes melhoramentos possíveis para esta abstracção: *a*) melhorar a heurística na criação da mesma - i.e., quando temos vídeos de notícias, uma boa *skim* seria a introdução do jornalista à notícia; e *b*) criar *skims* em tempo real, utilizando a informação das pesquisas.

Estas abstracções revelaram-se de extrema utilidade segundo os utilizadores finais. Existem pontos onde a interface, segundo estes, podia aumentar as suas funcionalidades - i.e, realizar pesquisas juntando múltiplos modos: imagem, texto, áudio.

Numa observação final sobre o projecto, foi sentido o problema de que esta biblioteca só podia ser utilizada com sucesso por aqueles que se encontravam na rede local, devido às taxas de transmissão necessárias.

Outra dificuldade encontrada está relacionada com os direitos intelectuais dos conteúdos. Apesar de existirem no projecto conteúdos públicos, muitos tinham um acesso restrito o que no final acabava por prejudicar a sua utilização.

2.1.3 MediaMill

O MediaMill [SSW10, SWG+04] é um sistema de pesquisa em arquivo de vídeo baseado em conceitos semânticos. Como arquivo foram utilizadas 184 horas de vídeo disponibilizadas pela TRECVID [tre10] em 2004, com conteúdos da ABC World News Tonight e CNN Headline News. Este sistema utiliza uma base de 32 conceitos[4] para catalogar o conteúdo do arquivo. Este processo da extracção semântica é dividido por três fases (Fig. 2.7):

Conteúdo

Nesta fase o vídeo é visto através de uma perspectiva de dados. É feita uma análise visual a cada 15 imagens do vídeo onde são extraídas características. O aúdio também é analisado e tido em conta com o léxico de palavras a que se pode associar conceitos - e.g., passageiro, carril, comboio, locomotiva são associadas ao conceito comboio. Os 32 conceitos são então ordenados pela sua probabilidade de ocorrência.

Estilo

O vídeo nesta fase é visto de uma perspectiva de produção, é analisado com base num conjunto de quatro aspectos:

disposição - duração, texto sobreposto, silêncio, voz-off

conteúdo - faces, localização da face, carros, movimentação do objecto, locutor frequente, extensão de texto sobreposto, entidade anunciada em texto, entidade anunciada em voz

captura - distância da câmara, trabalho da câmara, movimentação da câmara

contexto do conceito - repórter, classificação da fase de conteúdo

Contexto

A última fase é composta por duas configurações possíveis, através de uma junção dos resultados anteriores e com recurso a uma ontologia, para a detecção final dos conceitos.

[4]Lista de conceitos: avião a levantar voo; futebol americano; animal; baseball; cesto de basquetebol; praia; bicicleta; Bill Clinton; barco; prédio; carro; cartoon; notícia financeira; golfe; animação; hóquei no gelo; Madeleine Albright; notícias; assuntos de notícias em monólogo; espaços exteriores; texto sobreposto; pessoas; pessoas a andar; violência física; estrada; futebol; evento desportivo; mercado bolsista; estúdio; comboio; vegetação; notícia meteorológica

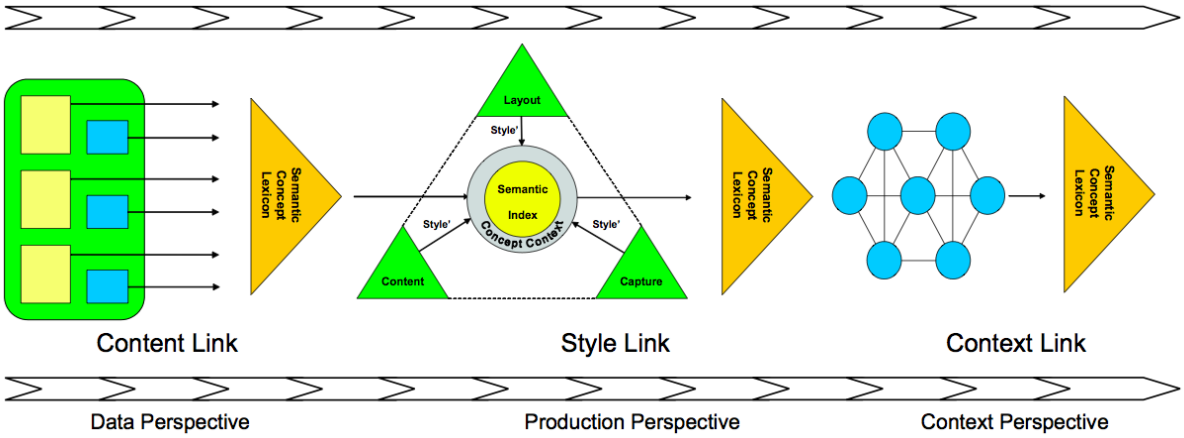
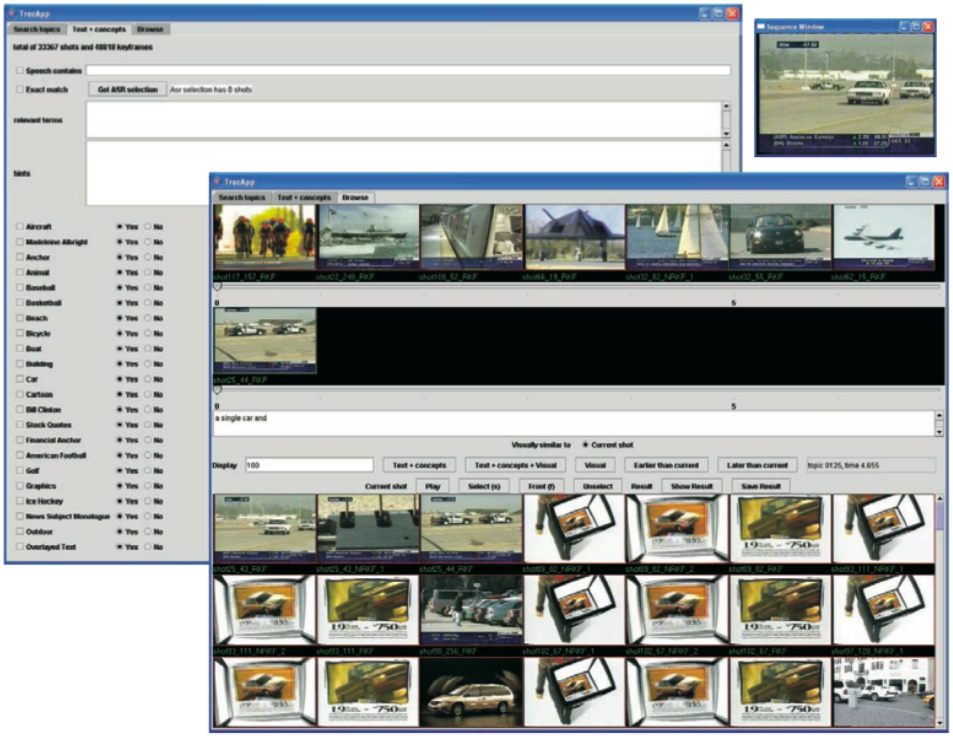


Figura 2.7: As três fases do processo de extracção semântica do MediaMill.

Ao nível da interface, foram utilizados dois modos (Fig. 2.8): um modo interactivo e um modo pessoal. No primeiro, o utilizador pode pesquisar por conceito ou por palavra-chave que foram retiradas pelo reconhecimento automático de voz. É possível também uma pesquisa por exemplo, onde para todas as *keyframes* é calculado o histograma de cada canal de cor e é realizada uma distância euclidiana para a comparação. A possibilidade de combinar conceitos na pesquisa torna possível por exemplo pesquisar por imagens de "Bill Clinton" e a "bandeira americana". O segundo tem por objectivo que o motor de pesquisa funcione mediante o perfil do utilizador que é aprendido com a sua utilização. Com recurso a uma ontologia baseada na WordNet [wor10], é criada uma estrutura sobre os conceitos para aumentar as possibilidades de pesquisa.

Nos testes efectuados pelos autores, é descrito que esta abordagem, de uma análise por fases, tem boas contribuições para determinados conceitos mas que para outros como "cesto de basquetebol", não existe uma melhoria de resultados. Também é mencionado o cuidado necessário a ter quando é feita uma atribuição semântica aos vídeos - e.g., para o conceito de "barco", foi definido que "vegetação" seria pouco provável de ocorrer, o que fez com que alguns barcos fossem muito comuns nos resultados e outros como caiaques em florestas eram praticamente eliminados do resultado final.

O maior desafio apontado pelos autores, será estender o léxico dos conceitos semânticos de forma a ser compatível com o conhecimento humano.



(a) Modo interactivo

(b) Modo VIPER (VIdeo PERsonalizer)

Figura 2.8: Dois protótipos para o sistema MediaMill

2.2 Segmentação de vídeo

Entende-se por segmentação de vídeo, a divisão em unidades do vídeo, segundo um determinado critério escolhido. A divisão mais comum nos vídeos é feita pelo conteúdo do vídeo de forma geral - i.e., através da criação de capítulos para uma história. Uma alternativa a esta, e com uma granularidade mais fina, é de fazer esta divisão pelos cortes de cena. Esta é a divisão que mais ajuda a tarefa de indexação dos vídeo, permitindo assim uma melhor representação do conteúdo do vídeo, visto o acesso eficiente a grande quantidades de informação não ser uma tarefa fácil. Na figura 2.9, podemos observar um sumário geral de um vídeo.

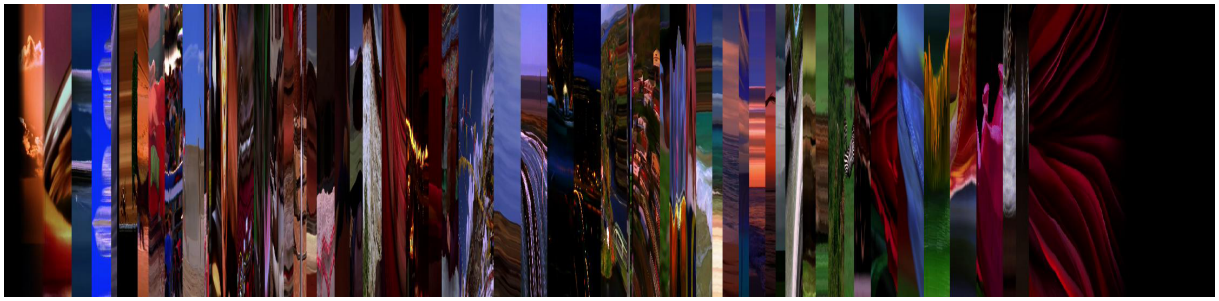


Figura 2.9: Sumário geral de um vídeo

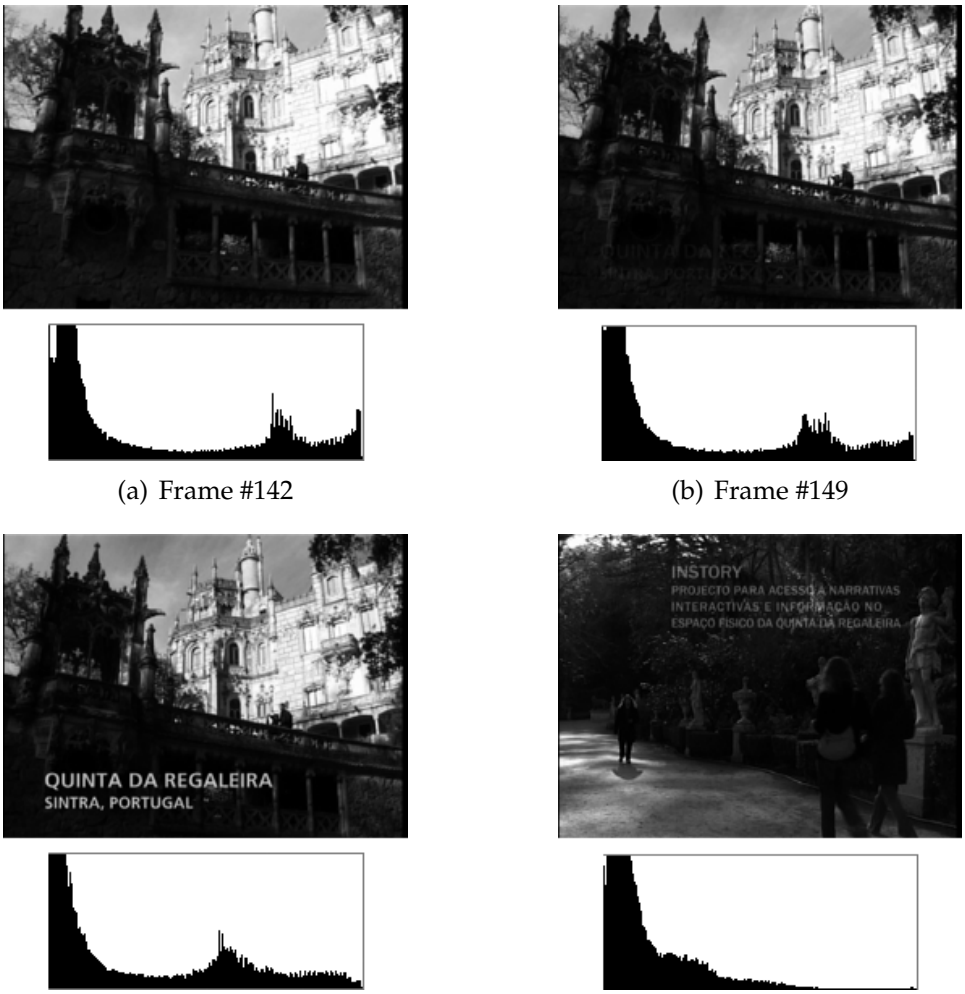
Esta imagem foi construída através da junção da coluna central de cada imagem do vídeo. Apesar de ser um método pouco preciso, visto só utilizar uma parte muito reduzida dos dados disponíveis, já é possível fazer uma identificação de cenas presentes no vídeo em questão. Uma cena é caracterizada por ter pequenas variações nas imagens a que lhe pertencem. A partir do momento em que se extraem os cortes de cena, podemos utilizar uma *frame* representativa da cada cena para a identificar e caracterizar, que servirá de base para futuras pesquisas.

No estudo de A. Dailianas, R. Allen e P. England, foi realizada uma comparação entre algumas técnicas de segmentação [DAE95], de onde se retiram várias conclusões. Dentro dos resultados obtidos, as que melhor se adequam às tarefas esperadas são: Diferença absoluta de histogramas (2.2.1) e Diferença ponderada de histogramas (2.2.2). O primeiro algoritmo é mais rápido de calcular e tem menos falsos positivos (56% por oposição dos 175%), no entanto o segundo tem melhores resultados (94% por oposição dos 73%). Uma vez que existe um alto número de falsos positivos, a validação humana continua a ser necessária. Isto não invalida a utilização destas técnicas, pois estas técnicas continuam a ser mais rápidas que uma segmentação manual. Uma técnica para eliminar alguns falsos positivos, nomeadamente em zonas com efeitos , como o *fade*, é

a introdução de um parâmetro de salto de imagens aquando da detecção de uma corte de cena, permite ao algoritmo eliminar a ocorrência destes falsos positivos.

2.2.1 Diferença absoluta de histogramas

Uma forma de observar estas variações é através do histograma de cada imagem (Fig. 2.10). Estes histogramas podem ser construídos com base na imagem em tons de cinzento, que tem uma representação mais simples que em RGB mas igualmente eficaz para a descrição necessária.



(a) Frame #142 (b) Frame #149

(c) Frame #160 (d) Frame #161

Figura 2.10: Amostra de *frames* e respectivos histogramas de um vídeo

Seguindo esta característica, podemos considerar que os objectos que se encontram na mesma cena irão permanecer, o que conduz a um histograma sem grandes variações. É possível observar este comportamento pelas figuras 2.10(a), 2.10(b) e 2.10(c), onde existem variações mínimas no histograma. Quando é feita a passagem da figura

2.10(c) para a figura 2.10(d), a variação já se torna significativa, o que indicará um corte de cena. No entanto, este valor de variação ou *threshold*, deve ser escolhido consoante as características do vídeo que estamos a analisar. Se for um vídeo em que a câmara que o gravou está fixa a apontar para um determinado local, então grande parte do vídeo final terá sempre a mesma imagem com maiores ou menores variações na mesma. Nestes casos é necessário escolher um valor mais baixo para o *threshold*, de modo a conseguir detectar as alterações. A segmentação é então calculada através da diferença absoluta dos histogramas, em que é somado o valor dos pixels de cada *frame* e feita a sua subtracção. Se esta for superior ao *threshold* indicado, então é adicionada à lista de cortes de cena. No final seleccionam-se as diferenças mais significativas de forma a encontrar as cenas mais relevantes.

2.2.2 Diferença ponderada de histogramas

Esta técnica utiliza os três canais de cores das imagens (Fig. 2.11), por oposição à técnica anterior. Visto uma imagem poder ter uma cor dominante, esta deverá ter um peso maior na comparação entre imagens.

$$d(f,f') = \frac{r}{s}d(f,f')^{(red)} + \frac{g}{s}d(f,f')^{(green)} + \frac{b}{s}d(f,f')^{(blue)}$$

Onde $d(f,f')$ é a diferença entre as imagens f e f'; r, g e b são as componentes de cada canal e s é calculado através $(r+g+b)/3$.

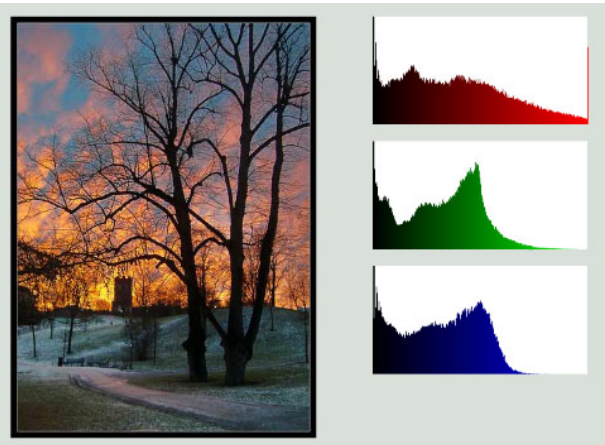


Figura 2.11: Exemplo de um histograma dividido pelos três canais de cor.

2.2.3 Intersecção de histogramas

Quando é realizada a intersecção de dois histogramas idênticos, o valor é máximo e igual ao número de *pixels* na imagem e menor quando existe um diferença. A semelhança entre imagens é dada por:

$$s(f,f') = \sum_{i=0}^{N} min(H(f,i),H(f',i))$$

2.2.4 Momentos invariantes

Os momentos invariantes têm propriedades como invariância à mudança de escala, rotação e transição o que os torna uma boa opção para a detecção de cenas. Esta técnica apresenta bons resultados quando usada em combinação com a técnica de intersecção de histogramas mas apresenta uma taxa de resultados correctos de apenas 54%. No entanto, mostra-se mais invariante às mudanças de *threshold*, o que poderá servir como uma segunda iteração numa segmentação de um vídeo.

2.2.5 Detecção de arestas

Este método propõe que uma mudança de cena possa ser detectada pela diferença dos locais onde as arestas aparecem. Esta técnica também utiliza um registo para calcular o movimento geral da imagem para ter em conta a percentagem de arestas que se encontram distantes das mais recentes. Apesar de melhores resultados, esta técnica tem um peso computacional muito superior em relação às outras técnicas.

2.2.6 Algoritmos genético para segmentação

Em [CGP+00] é descrita uma técnica de segmentação através de algoritmos genéticos. A técnica inicia-se com um pré-processamento automático onde é criado um conjunto representativo do vídeo com uma taxa de amostragem de duas imagens por segundo. Deste conjunto seleccionam-se as imagens mais distintas[5] através da diferença de histogramas descrita anteriormente, criando um subconjunto F'. Para este algoritmo é dado além do vídeo, um número k de cortes de cena esperados, retornando os k cortes e a sua pontuação de importância.

Os algoritmos genéticos definem uma população inicial e realizam uma série de operações de forma a tentar optimizar a próxima geração.

[5] São escolhidas as imagens cuja a diferença esteja σ acima da diferença média do conjunto.

Estes algoritmos são compostos por vários elementos[6], sendo que para o caso da segmentação, os mais importantes são os seguintes:

Função de *Fitness*

Esta função tem com objectivo a optimização da solução final, atribuindo uma posição a cada indivíduo na população. A função é definida pelas semelhanças adjacentes (*similarity adjacency functions*) que utiliza os valores das diferenças anteriormente referidas ou então atribui um peso a cada elemento do conjunto para a sua importância[7] no vídeo.

Operação de Mutação

Esta operação introduz diversidade de geração para geração da população através de trocas nos *bits* de codificação de cada indivíduo. Como cada *bit* da codificação de um indivíduo representa a presença ou ausência de corte, a mutação envolve a negação de um *bit*. Esta operação pode não beneficiar o algoritmo, visto existir a possibilidade de reduzir o número de cortes. Assim a mutação é apenas realizada para garantir que caso as operações de trocas gerem um maior ou menor número de segmentos esperados, seja feita a correcção para o valor k esperado num indivíduo.

Este tipo de algoritmos oferece resultados onde existe uma correlação entre a qualidade e o tempo de execução nos mesmos. Na figura 2.12, é exibido o resultado de uma segmentação para um número de cortes k=5, que segundo os autores obteve uma boa representação do vídeo base. Os autores argumentam também que a vantagem da utilização dos algoritmos genéticos é de se optimizar a função de *fitness* para se focar nas características desejáveis para executar a segmentação, mantendo-se o resto inalterável.

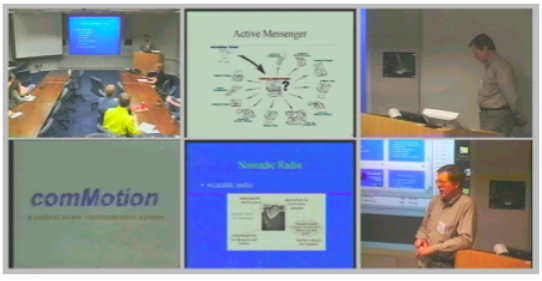


Figura 2.12: Resultado da aplicação do algoritmo genético para realizar uma segmentação. Imagem retirada de [CGP+00].

[6]Os elementos mais habituais são: Codificação, Função de Fitness, Operação de Troca, Operação de Mutação

[7]Um elemento menos comum e que ocorra num maior período de tempo, terá maior importância.

2.2.7 Aplicações da segmentação de vídeo

Como forma de reduzir o espaço necessário para guardar os metadados com redundância (ao nível de *keypoints*), a segmentação de vídeo é uma óptima solução. Tendo como pressuposto que uma cena tem uma duração de 1 segundo (algo que até é um pressuposto pouco optimista) como a *framerate* em média é de 25 FPS, conseguimos uma redução para 1/25 do espaço necessário.

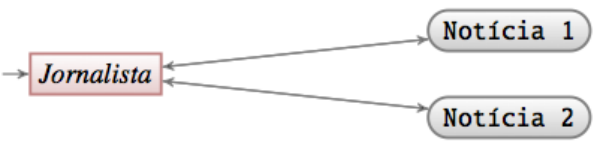
Através da criação de um grafo de cenas [YYL98], consegue-se criar um modelo de representação da acção do vídeo. A figura 2.13 mostra a sequência normal de cenas de um programa informativo. Este grafo pode ser particularmente útil para estruturar os conteúdos do vídeo, ajudando as pesquisas a direccionarem-se para as cenas relevantes. Neste caso, as pesquisas podem focar-se nas cenas de "Notícia", descartando as cenas onde aparece o jornalista.

(a) Jornalista (b) Notícia

Figura 2.13: Sequência de cenas de um programa de informação.

O grafo deste tipo de sequências será muito semelhante ao mostrado na figura 2.14(b).

(a) Uma sequência sem lógica associada.



(b) Grafo gerado através de uma comparação entre cenas.

Figura 2.14: Representação de cenas de um vídeo.

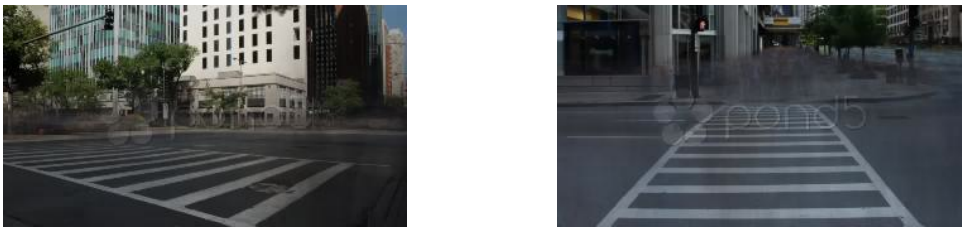
Este grafo é construído através de uma comparação de distâncias entre cenas permitindo assim chegar às cenas de conteúdo semelhante. Para obter este resultado existem vários processos possíveis:

(1) - Utilização da *frame* de identificação de cena

Os cálculos de diferenças entre cenas são feitos através das diferenças entre as *frames* que identificam cada cena - e.g., a primeira *frame* da cena.

(2) - Geração de imagem *background*

É gerada uma imagem *background* [CLL08] (Fig. 2.15) para cada cena, que é construída através da média de todas imagens que ocorrem na cena em questão. O cálculo é então realizado através da diferença destas imagens geradas para cada cena. Naturalmente este é um processo mais dispendioso computacionalmente. No entanto este processo consegue uma melhor representação do local onde decorre a cena, que não era possível com a extracção de uma única imagem, visto existirem objectos que poluíam a mesma.



(a) (b)

Figura 2.15: Exemplos de *background* de imagens. Com esta técnica é possível extrair o fundo de uma cena.

(3) - Combinação de ambos (1 e 2)

Juntar os dois processos anteriormente descritos e atribuir um peso a cada um deles.

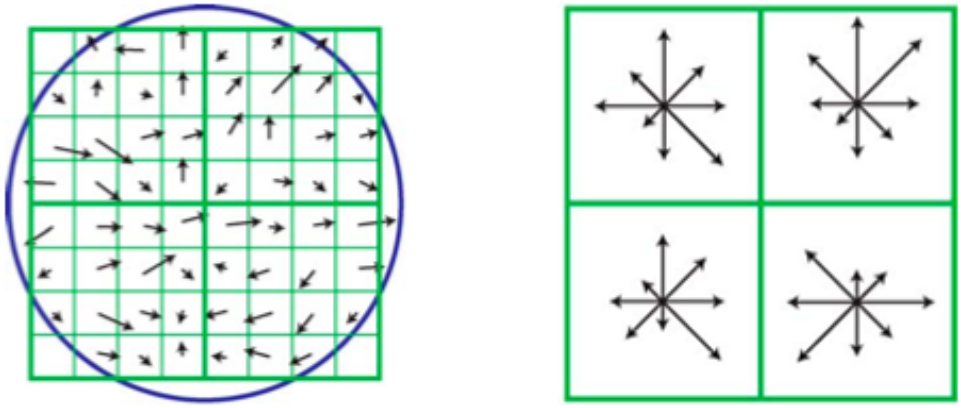
Deste modo é então possível gerar o grafo. Ao observar o grafo deste exemplo, verifica-se que existe um nó que se "repete"ao longo do vídeo. Assim um conjunto de cenas que antes eram independentes, podem ser relacionadas.

2.3 Análise de imagens

Nesta secção serão descritas técnicas para realizar análise às imagens, de forma a extrair metadados que possam ser utilizados posteriormente. Estas são as técnicas mais utilizadas na criação de descritores de imagens, detecção de faces e de conceitos.

2.3.1 Scale-Invariant Feature Transform (SIFT)

O método SIFT [Low04] é caracterizado por duas fases: detecção de pontos de interesse (*keypoints*) e extracção do descritor visual. O seu objectivo é de encontrar *keypoints* que representam locais relevantes da imagem e que se mantenham estáveis em relação às mudanças de escala e rotação. São utilizados filtros de diferenças Gaussianas para detectar estes *keypoints*.

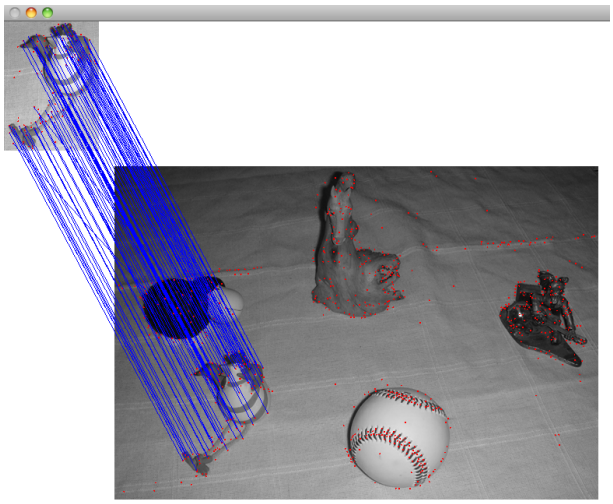


(a) Gradiente (b) Descritor

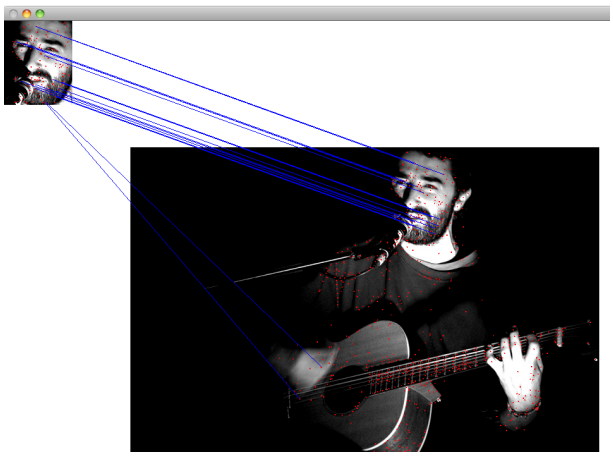
Figura 2.16: Exemplo do descritor numa região 8x8. Esta imagem foi retirada de [Low04].

No primeiro passo, é obtido o gradiente (Fig. 2.16(a)) de cada ponto numa região de 16x16[8]. De seguida para cada bloco de 4x4, é calculado o histograma com 8 direcções do gradiente. Cada uma destas regiões é representada por 128 valores[9]. O descritor (Fig. 2.16(b)), devido às propriedades de cada *keypoint*, é caracterizado por ter invariância à iluminação, rotação e escala. Este descritores apresentam um grau de distinção que permite que funcionem como identificador da imagem em questão. Através dos testes realizados até este momento, é possível identificar que tipos de imagens podem funcionar melhor como elemento de pesquisa (Fig. 2.17).

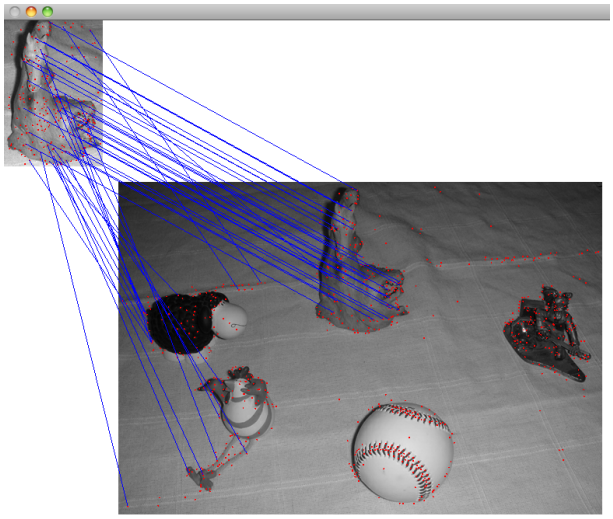
[8]Valor por omissão, apesar de existir a opção de alterar o tamanho da região.
[9]Este valor é obtido através de 16 regiões 4x4 vezes 8 pontos do histograma.



(a)



(b)



(c)

Figura 2.17: Exemplos com um elevado grau de precisão. Nestas imagens os pontos vermelhos representam os *keypoints* identificados pelo algoritmo e as linhas azuis unem dois *keypoints* que foram considerados semelhantes pela função de comparação.

Uma imagem com pouca variação de cores irá criar poucos *keypoints*, o que irá afectar o resultado final visto existirem menos elementos de comparação, o que pode criar falsos positivos (Fig. 2.18).

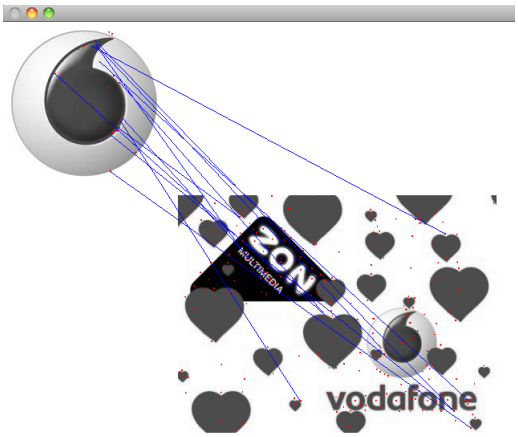


Figura 2.18: Falsos positivos.

Este algoritmo obteve bons resultados, que são explicados pela captura de grandes quantidades de informação relacionada com padrões de intensidade espacial, sendo ao mesmo tempo robusto para pequenas deformações ou erros localizados. No entanto o *matching* pode ser melhorado introduzindo mais parâmetros - e.g., a localização dos *keypoints*, de forma a que nos exemplos da figura 2.17(b) e 2.17(c), seja possível eliminar aqueles que se encontram fora da área esperada.

2.3.2 Speeded Up Robust Features (SURF)

O algoritmo SURF [BTG06] é uma técnica semelhante ao SIFT para a criação de descritores de imagens. Nesta técnica, nem o detector de pontos, nem o descritor utiliza dados relacionados com a cor. Na detecção de pontos é utilizado o método de Imagens Integrais (*Integral Images*) que reduz drasticamente o tempo de computação. Este método permite a computação rápida de filtros de convolução aplicados a determinadas áreas. Assim o valor para um determinado ponto é a soma de todos os pixels. São depois necessários mais três somas para calcular o valor para a área desejada (Fig. 2.19)[10]. O detector de pontos é baseado na Matriz Hessian devido à sua precisão. São detectadas estruturas de *blob*[11], onde o determinante na matriz é máximo. Também se

[10]Todas as imagens desta secção foram retiradas de [BTG06]

[11]Um objecto binário de grandes dimensões, também conhecido como *blob*, é uma colecção de dados binários armazenados como uma única entidade num sistema de base de dados.

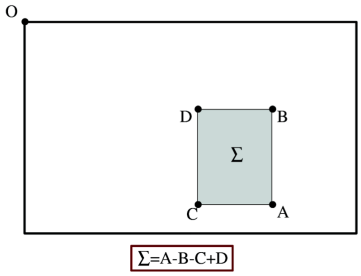


Figura 2.19: Exemplo de uma imagem integral

recorre ao determinante para a escolha da escala. Dado um ponto x = (x,y) de uma imagem a matriz para a escala σ é definida

$$\mathscr{H}(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}$$

onde, $L_{xx}(x,\sigma)$ é a aplicação da derivada de segunda ordem Gaussiana, visto serem óptimas para analise escala-espaço.

Para localizar os pontos de interesse na imagem para várias escalas, é feita uma supressão de valores não máximos num região de 3 x 3 x 3. O máximo do determinante é então interpolado na escala e espaço da imagem. Esta interpolação é importante, visto a diferença entre as primeiras escalas ser muito significativa (Fig. 2.20)

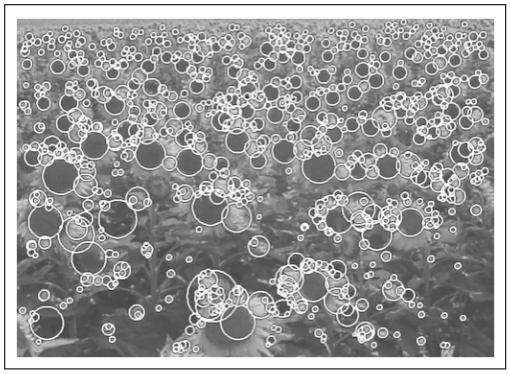


Figura 2.20: Detecção de pontos de interesse

Os descritores do SURF utilizam apenas 64 dimensões em oposição dos descritores do SIFT e são construídos através da transformada de Haar, de forma a tirar partido das imagens integrais.

O primeiro passo para a extracção do descritor é selecção de uma região centrada no ponto de interesse. A região é divida em partes de 4 x 4 onde são aplicadas as

transformadas de Haar. Os resultados de cada sub-região são somados, criando um vector de características. Visto cada sub-região ter um vector de quatro dimensões, a concatenação de todos, produz um vector de 64 dimensões (Fig. 2.21).

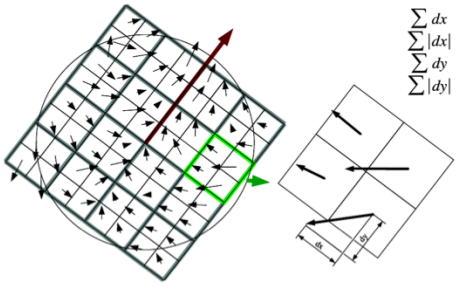


Figura 2.21: Descritor SURF

Uma das características é a invariância relacionada com o contraste, que é conseguida através do factor da escala. Esta invariância será muito importante aquando da comparação de descritores.

Foram realizados testes pelos autores, para identificar a presença de objectos em imagens, comparando os resultados com testes realizados com o algoritmo SIFT e uma terceira técnica (Fig. 2.22). Com uma amostra de 400 imagens, foi possível verificar que o SURF consegue resultados ligeiramente superiores aos alcançados pelo SIFT.

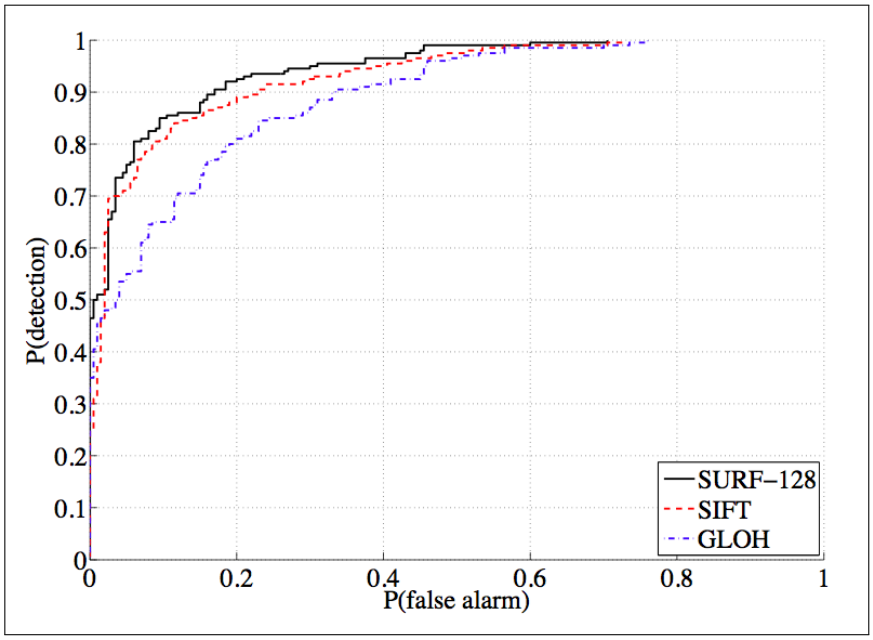
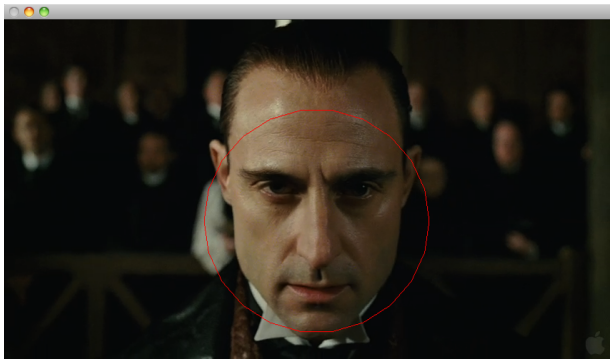


Figura 2.22: Comparação dos resultados entre três técnicas

2.3.3 Detecção de faces

O trabalho de Paul Viola e Michael Jones [VJ01] permite, através de uma cascata de classificadores, previamente treinados, identificar objectos numa imagem em tempo real. O processo começa com a cascata contendo apenas um classificador. Com o desenrolar do processo a complexidade vai aumentando com a junção de outros classificadores, até chegarmos à confirmação de presença de faces em sub-janelas da imagem.



(a)

(b)

Figura 2.23: Exemplos da detecção de faces

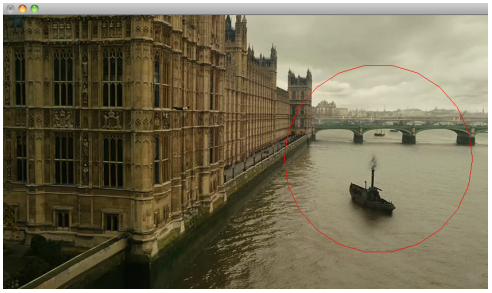
Nos testes realizados (alguns exemplos na figura 2.23), a detecção funciona com um elevado grau de precisão nos resultados obtidos, apesar de existirem:

(1) - Falsos Positivos

Normalmente, a zona identificada tem algumas semelhanças com as características de uma face (Fig. 2.24(a))

(2) - Falsos Negativos

O algoritmo não está preparado para detectar faces de perfil (Fig. 2.24(b)), visto o classificador não conter este tipo de faces quando foi treinado. A face tem dimensões muito reduzidas.(Fig. 2.24(c))



(a)

(b)

(c)

Figura 2.24: Exemplos de falhas na utilização

Após a extracção, é possível fazer uma classificação das faces de modo a tentar identificar pessoas. Em [GJC09] foram utilizadas, sobre os resultados provenientes do processo anteriormente descrito, técnicas para identificar a pele, a pose e ainda o género (Fig. 2.25). Com esta segunda análise sobre os dados, é possível reduzir a quantidade de falsos positivos, pelo resultados indicados no estudo. É ainda adicionado um nível adicional de informação às faces, como é o caso do género.

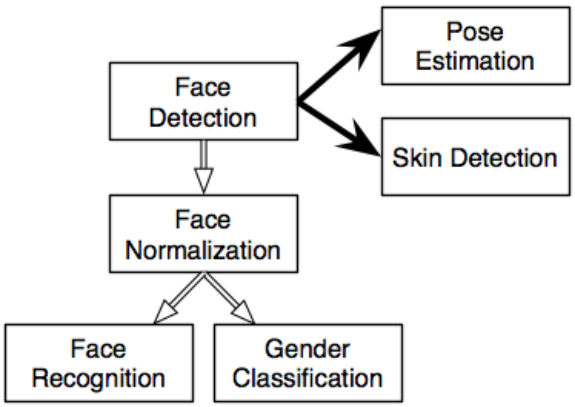


Figura 2.25: Detecção de pele e pose em faces

2.3.4 Conceitos semânticos

Como forma de retirar mais metadados dos vídeos do arquivo, a possibilidade de detectar conceitos existentes nas cenas do vídeo - e.g., cenas com animais, veículos - aumentará as possibilidades de pesquisa. Através da detecção de conceitos em cenas, será possível criar processos para atribuir anotações automáticas.

Nesta secção será descrita a técnica para a detecção de conceitos (secção 2.3.4.1), assim como ontologias observadas (secção 2.3.4.2) que ajudam na estruturação dos resultados.

2.3.4.1 Técnica para detecção de conceitos

A proposta descrita em [Jes09], é baseada em classificadores Regularized Least Squares (RLS). São realizadas classificações binárias - i.e., conjuntos positivos e negativos (Espaço Interior vs Espaço Exterior ou Com Pessoas vs Sem Pessoas) - para o conjunto de imagens e é utilizado uma função sigmóide para converter o resultado do classificador numa pseudo-probabilidade.

Para construir um classificador capaz de detectar um conceito, é necessário treiná-lo previamente com um conjunto de teste (subconjunto de imagens com o conceito e outro subconjunto sem o conceito), de onde são extraídas as características que serão tidas em conta quando da classificação de uma nova imagem. Uma dessas características é a cor, assim são utilizados momentos de cor no espaço Hue Saturation Value (HSV). Este espaço, pelas suas características de invariância, é indicado para analisar a cor. A imagem é dividida em 9 blocos onde a média e variância de cada canal de cor é observado. No entanto, esta divisão pode introduzir descontinuidades em objectos, assim é feita uma segmentação utilizando o algoritmo Mean-Shift (Fig. 2.26)[12].

(a) Imagem Original (b) Imagem Segmentada

Figura 2.26: Regiões de cor utilizando o algoritmo Mean Shift.

[12]Todas as imagens desta secção foram retiradas de [Jes09].

É também aplicado um banco de filtros de Gabor para a detecção de texturas (Fig. 2.27).

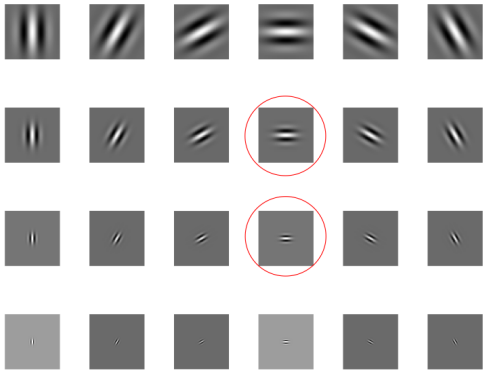


Figura 2.27: Banco de filtros Gabor

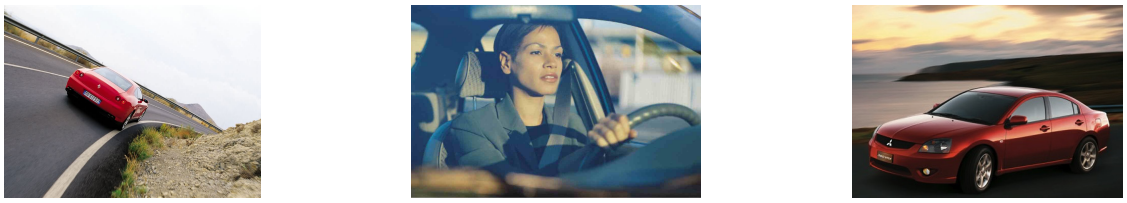
Foram treinados classificadores para os seguintes conceitos: Espaço Interior, Neve, Praia, Natureza, Face, Festa, Pessoas. Como é esperado, quanto mais abstracto for o conceito, menor será a taxa de sucesso na detecção do mesmo, visto ser mais difícil encontrar características que o possam representar. Deste modo, conceitos muito genéricos podem introduzir um elevado número de falsos positivos ou falsos negativos no resultado final.

2.3.4.2 Ontologias

Para a escolha de conceitos, uma primeira abordagem, assim como uma boa prática para o estudo de quais os conceitos a detectar, será a construção de uma ontologia. Uma ontologia é construída para representar e organizar os elementos a ela pertencentes e também para representar relações que possam existir entre os mesmos. Quanto melhor for a descrição do domínio, menor será a ambiguidade da linguagem livre. Passa a existir uma forma unívoca para identificar o elemento, neste caso o conceito [NST+06].

Na detecção de conceitos, apesar de haver grandes avanços - como é possível observar em [SS10] - a quantidade de conceitos detectados continua a ser significativamente pequena comparada com o léxico humano. As ontologias podem contribuir para um aumento desse léxico. Se o utilizador desejar construir um classificador para detectar o conceito de "peão" - i.e., uma pessoa na rodovia - pode juntar várias imagens com peões para o conjunto positivo e outras que não os contenham para o conjunto negativo. No entanto, o ambiente possível onde está inserido o peão é tão diverso que o

classificador no final terá forte probabilidade gerar falsos positivos ou falsos negativos. Uma particularidade das ontologias, está no facto de com regras conseguirmos um aumento da semântica, através de inferência. Com a combinação de conceitos que já foram alvo de muita investigação, como é o caso de "pessoa", "carro" ou "estrada", podemos reutilizá-los e aquando da análise da imagem para a detecção de conceitos inferir que, se existir "pessoa" e "carro" na imagem ou 'pessoa" e "estrada", existe uma forte probabilidade dessa pessoa ser um peão. Desta forma, reduzimos o número de classificadores treinados e tentamos extrair essa informação através de conceitos mais fáceis de detectar. Anteriormente foi descrita a possibilidade de reconhecer se uma pessoa pode ser um peão através de outros conceitos, mas se esses conceitos ocorrerem em cenas diferentes do vídeo, só com a ajuda de ontologias e regras é possível tirar partido do factor temporal existente no vídeo. As cenas na figura 2.28, são planos comuns em sequências de filmagem profissional, onde o plano do condutor é inserido entre cenas do veículo para enquadrar o assunto.



(a) Cena A - carro (b) Cena B - face (c) Cena C - carro

Figura 2.28: Exemplo da junção de conceitos, ontologias e regras.

Desta forma, a criação de regras sobre as ontologias, pode tirar proveito de regras/-planos/sequências cinematográficas para aumentar a semântica extraída.

Para aumentar a capacidade das regras, estas podem também utilizar informação extraída do som além daquela proveniente da imagem - e.g., o som do público de uma palestra é muito diferente do público de um estádio de futebol, apesar de ambas detectarem os conceitos de "faces" ou "pessoas".

Descrevendo uma janela de observação, a regra pode utilizar um conjunto de cenas anteriores e posteriores e com conceitos existentes nestes conseguir inferir novos conceitos, ou conceitos mais concretos para a cena corrente.

A Large Scale Ontology For Multimedia (LSCOM) [NCHS10] é uma ontologia que foi construída com base em milhares de termos usados em pesquisas pela BBC (British Broadcasting Corporation) nos finais dos anos 90 em conjunto com tópicos de pesquisa avaliados nas TREC Video Retrieval Evaluation (TRECVID) [tre10] de 2003 e 2004. As TRECVID são um conjunto de conferências, patrocinadas pelo National Institute of Standards and Technology (NIST), com o objectivo encorajar o estudo na recuperação de informação de vídeos, disponibilizando colecções de vídeo, procedimentos uniforme para resultados e fóruns de discussão.

A primeira versão da LSCOM é composta por 856 conceitos, relacionados com eventos, objectos, locais, pessoas e programas. A TRECVID de 2005 começou a utilizar uma versão reduzida desta ontologia, como base para quais os conceitos a serem avaliados (Fig. 2.29). Cada conceito foi avaliado em critérios de:

- **Utilidade** ou relevância em pesquisas a efectuar
- **Abrangência** ou cobertura do interesse do utilizador dentro domínio semântico
- **Observação** ou ocorrência do próprio nos vídeos.
- **Viabilidade de Detecção** numa perspectiva de ser possível a sua detecção através de técnicas, que caso não existam, possam emergir num espaço de 5 anos.

Este processo de avaliação iterativo envolveu solicitações de introdução de conceitos, a consulta especialistas de domínio e comparação com outras ontologias.

Este projecto tem o objectivo de atingir 1,000 conceitos e para tal estão a utilizar o repositório da Cyc Knowledge Base [cyc10], com o qual estão a fazer o mapeamento de conceitos. Este mapeamento irá ajudar a ontologia LSCOM a enriquecer a representação de conhecimento através de regras, assim como suporte para níveis semânticos não alcançados - e.g., o conceito de "Bandeira Americana" encontrava-se dentro de "Objectos" mas passou para dentro de um novo subconjunto "Bandeira" pertencente a "Objectos", de forma a incorporar futuras bandeiras.

Um outro factor será a produção de uma Web Ontology Language (OWL) do subconjunto relevante da Cyc que contém relações binárias - e.g., espaço interior e espaço exterior - assim como relações de alto-nível (*rule macros*) para a criação de regras que relacionam conceitos a relações binária - e.g., uma pessoa ser considerada peão, se existir também um veículo ou uma estrada.

Figura 2.29: Organização da LSCOM (Versão Lite).

O EUROVOC - Thesaurus [eur10a] é uma ferramenta multilíngue mantida pelo Serviço de Publicações da União Europeia [Eur10b] com o objectivo de cobrir o domínio da actividade das Comunidades Europeias. Uma vantagem desta estrutura é que permite aos utilizadores interrogarem o sistema documental na sua própria língua visto ser independente da língua de indexação. Este tesauro (anexo A), está em constante actualização e é estruturado da seguinte forma:

Domínios

Designação mais abstracta, identificada por dois algarismos - e.g., 28 - Questões Sociais.

Micro-tesauros

Sub-categoria de um domínio, identificada por quatro algarismos, onde os dois primeiros pertencem ao domínio - e.g., 2826 - Vida Social.

Descritores

Palavras ou expressões que descrevem de forma não ambígua os conceitos que constituem o domínio - e.g., Tempos Livres.

Anotações

As anotações têm por objectivo ajudar à definição do descritor ou na aplicação do descritor aquando da sua indexação aos elementos - e.g., desportos, férias, turismo.

A Duvideo está a utilizar este tesauro para catalogar todo o seu arquivo de vídeo. Desta forma, foi realizado um mapeamento (anexo B) entre o EUROVOC e uma lista de 100 conceitos utilizados pela ImageCLEF [ima10] - onde aqueles enunciados na tabela 2.1 foram mapeados directamente - que incorpora os conceitos da LSCOM Lite. Os outros conceitos como "peão" chegam através de uma regra de composição como foi explicado anteriormente (Fig. 2.30).

Tabela 2.1: Conceitos retirados da ImageCLEF para mapear com o EUROVOC

Flores	Plantas	Árvores	Montanha	Água
Peixe	Deserto	Instrumento Musical	Carro	Avião
Veículo	Bicicleta	Comboio	Barco	Brinquedo
Insecto	Ave	Lago	Mar	Bebé
Jovem	Criança	Adulto	Idoso	Grupo de Pessoas
Igreja	Artes Visuais	Graffiti	Pintura	Animais
Cavalos	Praia	Desportos		

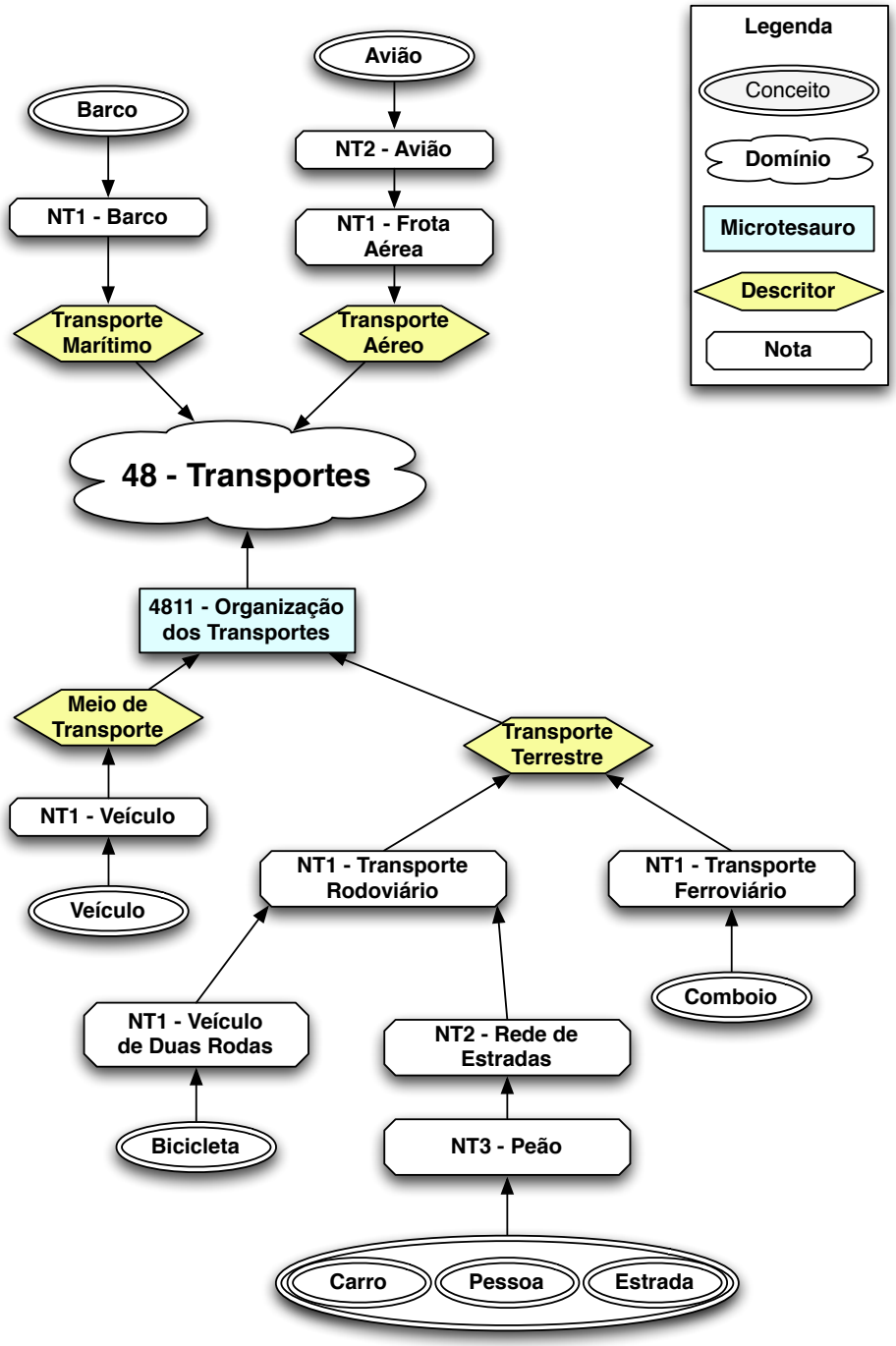


Figura 2.30: Exemplo de mapeamento entre o domínio "Transportes" do EUROVOC e conceitos.

2.4 Integração de metadados

Os metadados permitem uma utilização eficiente e eficaz dos recursos disponíveis [WC02, Wen99], ou seja dos vídeos. Estes podem ser categorizados em:

Descritivos
Facilitando a identificação de recursos, normalmente através da atribuição de categorias. Metadados descritivos ajudam na pesquisa e identificação de recursos - e.g., título, autor. Este tipo de metadados costuma ser público e a sua partilha encorajada.

Administrativos
Ajudando na gestão de recursos - e.g., o registo de quem visualizou um vídeo; se os metadados do vídeo já foram validados; permissões de acesso.

Estruturais
Este tipo de metadados são utilizados para estabelecer ligações de componentes de informação mais complexas dos recursos - e.g., a face extraída de uma cena do vídeo é identificada numa cena de um outro vídeo.

Assim a informação gerada através dos métodos enunciados nas secções anteriores (2.2 e 2.3) será incorporada nos vídeos como forma de conseguir atribuir dados semânticos relativos ao conteúdo dos vídeos. Para realizar esta integração, os formatos que permitem uma maior relação entre a Essência[13] e os Metadados são: Multimedia Content Description Interface (MPEG-7), Advanced Authoring Format (AAF) e Material eXchange Format(MXF). As secções seguintes fornecem as descrições de cada um destes formatos.

2.4.1 Multimedia Content Description Interface (MPEG-7)

O MPEG-7 é uma norma para a descrição de conteúdos multimédia com o objectivo de tornar a pesquisa e filtragem de conteúdos mais eficiente [Cab06, SS02]. A norma é composta por componentes que disponibilizam ferramentas para a descrição dos conteúdos. Este formato pretende atingir um elevado nível de interoperabilidade entre quem cria e utiliza os metadados, utilizando como base o XML. Existem vários níveis de abstracção quando é efectuada a extracção de características e informação técnica, sendo possível descrever desde a cor e textura ao local e data da criação do vídeo.

[13] Essência é o material relevante, que neste caso será o material audiovisual.

2.4.2 Advanced Authoring Format (AAF)

Este formato foi desenvolvido pelo Advanced Media Workflow Association (AMWA) (antiga AAF Association) e é um contentor (*wrapper*) da Essência e Metadados associados, direccionado para a pós-produção [San06]. Na figura 2.31, podemos observar a organização de um AAF.

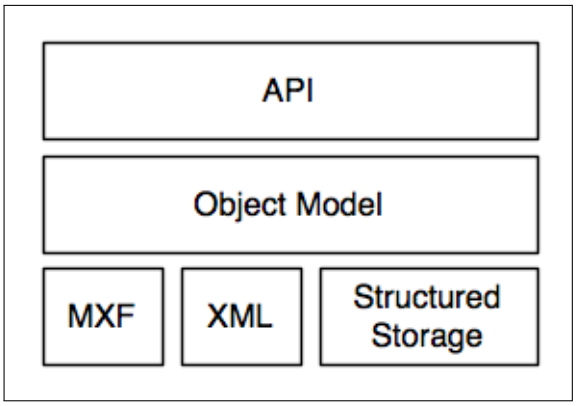


Figura 2.31: Organização de um ficheiro AAF

O *Object Model* contém os principais metadados que especificam a estrutura do conteúdo assim como informação já direccionada para a pós-produção como o número de faixas, *timecode* e compressão da Essência. Na camada de armazenamento, temos a possibilidade de utilizar XML (eXtensible Markup Language), MXF ou Structured Storage (um formato da Microsoft para envolver Essência e Metadados). Uma das características interessantes deste formato, é que é possível referenciar material que se encontra alojado externamente ao ficheiro AAF.

2.4.3 Material eXchange Format(MXF)

O formato MXF é um contentor para vídeo e áudio profissional, criado através de um conjunto de parâmetros definidos pela Society of Motion Picture and Television Engineers (SMPTE) [SMP09]. A sua uniformização torna-o num formato mais direccionado para um produto final ou muito perto de tal. O MXF é um formato mais simples que o AAF. Ao contrário do AAF, o MXF só consegue referenciar conteúdos que se encontram dentro do mesmo, ou seja, na sua Essência. Na concepção do MXF, a Professional-MPEG Forum (Pro-MPEG) e a AMWA acordaram um Zero Divergence Doctrine (ZDD) para que existisse um alinhamento com o modelo de dados do AAF. O MXF disponibiliza um método de troca de metadados que se torna independente do formato audiovisual o que é ideal para armazenar a informação associada [DWBT06].

No caso de ser necessário adicionar metadados para pós-produção, é possível adicionar extensões ao modelo de metadados [mog05].

Através da figura 2.32 é possível observar que o formato MXF é na realidade um conjunto de normas. Isto permite que o MXF se adapte às necessidades de cada tarefa, simplificando ao máximo a sua composição.

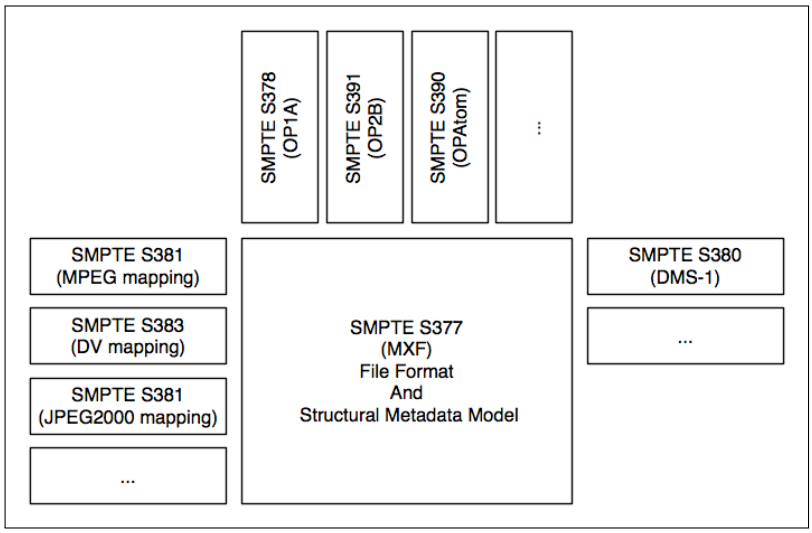
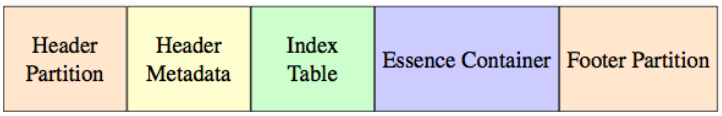
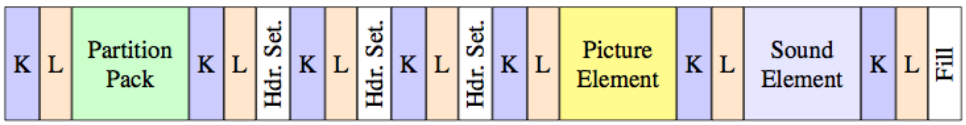


Figura 2.32: Arquitectura normalizada do documento MXF

Na figura 2.33, é possível observar a organização base de um ficheiro MXF. Como é possível verificar toda a informação se encontra num formato Key-Length-Value (KLV)[14] o que permite a descodificadores de MXF e a motores de processamento ignorarem informação que não compreendem - e.g., *keys* que não são reconhecidas.



(a) Vista de alto nível



(b) Vista de baixo nível

Figura 2.33: Estrutura do MXF.

[14] K - chave única para identificar o campo (16*bytes*); L - tamanho do campo; V - valor do campo

Os primeiros dois blocos (*Header Partition* e *Header Metadata*) representam o cabeçalho do ficheiro. A *Index Table* referencia todos os conteúdos que existem no *Essence Container*, de forma a facilitar o seu acesso.

A parte *Header Metadata* é o local de onde chegam os maiores benefícios ao MXF [Dev02]. É a área onde são adicionados os metadados e os parâmetros de tempo e sincronização são definidos.

A sincronização e a descrição da Essência é controlada por 3 módulos:

Material Package

Representa a *timeline* final do ficheiro pós-edição.

File Package

A Essência é descrita neste módulo. Esta encontra-se na sua forma original sem qualquer edição.

Source Package

Este módulo inclui todas as edições (Edit Decision List - EDL) efectuadas sobre a Essência.

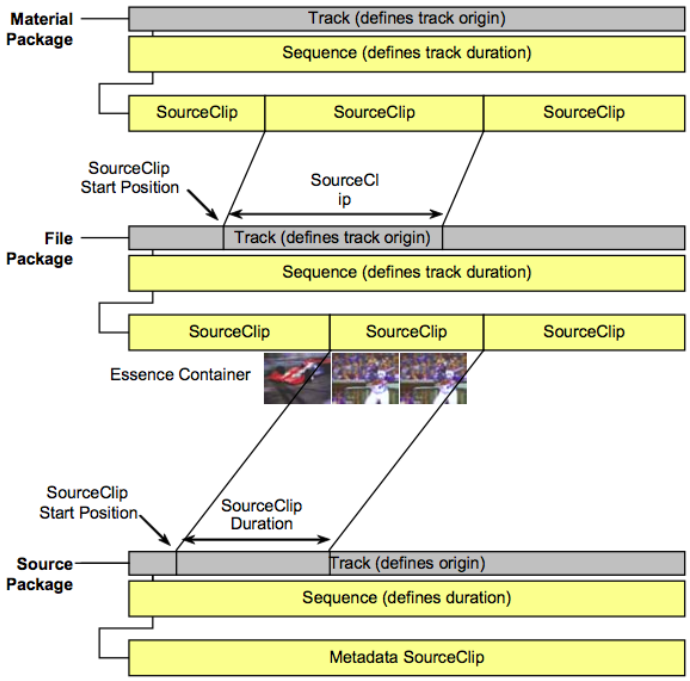


Figura 2.34: Ligação dos módulos

Se existir um único vídeo no Material Package, este irá corresponder ao File Package na sua totalidade, e não se tira partido desta representação. Quando existe mais do que um vídeo no ficheiro, toda esta organização se torna natural.

2.4.4 Extensão ao *schema* dos metadados no MXF

O *schema* nos MXF representa dois tipos de metadados (Fig. 2.35): Estruturais e Descritivos. É permitida uma alteração completa dos mesmo, no entanto isso traria os

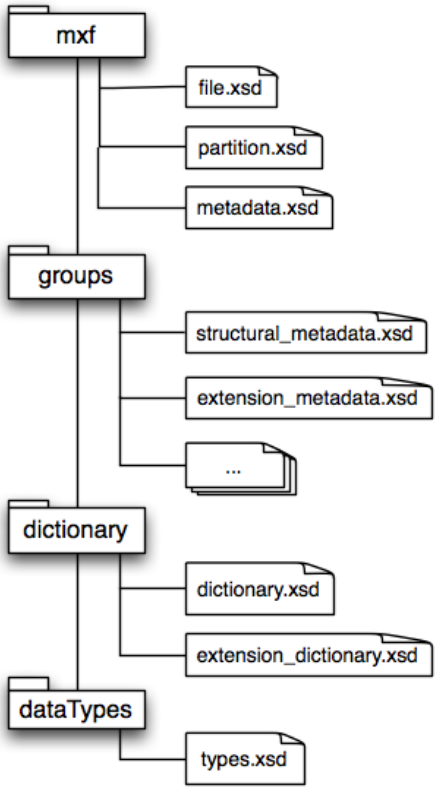


Figura 2.35: Estrutura dos *schemas*

problemas associados de perda de interoperabilidade com outras aplicações, por isso uma extensão é a melhor abordagem. O schema deve obedecer ao modelo dos MXF onde propriedades - e.g., face, conceito - tem de se encontrarem dentro dos *Metadata Sets* que são listas de propriedades - e.g., cena.

Cada propriedade deve ser adicionada ao ficheiro **dictionary.xsd** (ver listagem 2.1). Os *Metadata Sets* encontram-se definidos no ficheiro **metadata.xsd**. Estes, como são grupos de propriedades, usam referências para enunciar as propriedades (ver listagem 2.2).

Quando os documentos XML incluem informação relacionada com a estrutura física do ficheiro - e.g. *Index Table*, *Partition Packs* - estes devem ser validados com o ficheiro **file.xsd**.

Listagem 2.1: Exemplo de extensão - propriedade.

```
<!-- - -->
<xs:element name="CompanyName" type="types:UTF16CharString">
  <xs:annotation>
    <xs:documentation>Manufacturer of the equipment or application
    that created or modified the file</xs:documentation>
    <xs:appinfo source="urn:SMPTE:MXF:XMLInterface:Dictionary:ElementName">
    Company Name</xs:appinfo>
    <xs:appinfo
      source="urn:SMPTE:MXF:XMLInterface:Dictionary:ElementUL">
      06.0E.2B.34.01.01.01.02.05.20.07.01.02.01.00.00
    </xs:appinfo>
    <xs:appinfo source="urn:SMPTE:MXF:XMLInterface:Dictionary:Length">
    var</xs:appinfo>
    <xs:appinfo source="urn:SMPTE:RP210:DataElementDefinition">
      [RP210 Specifies the name of the application provider]</xs:appinfo>
  </xs:annotation>
</xs:element>
<!-- - -->
```

Listagem 2.2: Exemplo de extensão - *Metadata Set*.

```
<!-- - -->
<xs:element name="Preface" type="groups:PrefaceType">
  <xs:annotation>
    <xs:documentation>Defines the Preface set</xs:documentation>
    <xs:appinfo source="urn:SMPTE:MXF:XMLInterface:Groups:Name">
    Preface</xs:appinfo>
    <xs:appinfo
      source="urn:SMPTE:MXF:XMLInterface:Groups:Key">
      06.0E.2B.34.02.53.01.01.0D.01.01.01.01.01.2F.00
    </xs:appinfo>
  </xs:annotation>
</xs:element>
<xs:complexType name="PrefaceType">
  ..<xs:sequence>
    ..<xs:element ref="LastModifiedDate"/>
    ..<xs:element ref="DMSchemes"/>
    ..</xs:sequence>
    </xs:extension>
    </xs:complexContent>
</xs:complexType>
<!-- - -->
```

2.4.5 Produção

O MXF é o formato mais comum [San06] logo desde o início de todo o processo da criação de conteúdo. As câmaras produzem ficheiros MXF na captura de vídeo. No caso da Duvideo, a gravação é realizada a partir de uma XDCAM que faz captura do vídeo. Na estrutura utilizada pela câmara é criado, entre outros, o clip MXF assim como um ficheiro XML. Este ficheiro XML é uma cópia dos metadados gerados que se encontram dentro do MXF. Os metadados gerados automaticamente pela câmara são parâmetros técnicos (Estruturais e Adminstrativos) incluindo: Data; Identificador de clip (UMID); Descrição da câmara; Duração do *clip*; *Frame rate* e Número de canais de áudio. Além destes metadados, a câmara adiciona outros já relacionados com o conteúdo do vídeo (Descritivos). Se existir uma alteração na fonte de luz (ambiente, fluorescente), é colocada uma marca temporal a indicar esta mudança. Esta é uma informação que pode ser utilizada como complemento à segmentação ou como forma de ajudar a sua validação.

2.5 Resumo

Foram estudados dois sistemas de arquivo de vídeo (VideoSTAR e TIP) de modo a retirar experiências sobre o seu desenvolvimento e a incorporar os aspectos positivos nesta solução proposta. Para a biblioteca de algoritmos que este sistema irá necessitar de forma a realizar as tarefas pretendidas, foram apresentados: três algoritmos de segmentação de vídeo (Diferença absoluta de histogramas, Diferença ponderada de histogramas e Algoritmos genéticos) e quatro técnicas de análise de imagem (SIFT e SURF para criação de descritores visuais, Viola Jones para detecção de faces e detecção de conceitos). Todos os metadados gerados pelo cálculo destes algoritmos necessitam de ser integrados com os conteúdos. Assim foram analisados os três formatos de vídeo profissional mais utilizados: MPEG-7, AAF e MXF. Foi dada uma maior importância ao suporte MXF, visto este ser aquele que é utilizado no *workflow* da Duvideo, empresa que participa no projecto VideoFlow, no âmbito do qual foi realizada esta dissertação.

3

Solução

Neste capítulo será descrito o trabalho realizado, com base no estudo mencionado no capítulo anterior.

Existe um conjunto de requisitos para sistemas de arquivo de vídeo, de forma a melhorar a sua utilização, entre os quais os seguintes que são aqueles mais focados nesta solução:

- Segmentação de vídeo;
- Detecção de conceitos;
- Detecção e identificação de pessoas;
- Pesquisa com base em imagem;
- Integração dos metadados nos ficheiros de vídeo.

De modo a modelar estes requisitos, a secção Desenho (3.1) contém os diagramas necessários para a compreensão do sistema, ao nível das funcionalidades e das arquitecturas propostas para a organização do mesmo. Na secção sobre a Realização (3.2), é descrito como foi implementada a solução do protótipo, que foi dividido em duas componentes: Servidor e Cliente.

Por fim, são apresentados os resultados dos testes realizados ao protótipo, assim como a análise aos inquéritos realizados à utilização do mesmo. Estes dados serão descritos na secção de Avaliação (3.3).

3.1 Desenho

Nesta secção será mostrado o modelo de *use cases* (secção 3.1.1), arquitectura proposta para o sistema final (secção 3.1.2), assim como a arquitectura do ViewProcFlow (secção 3.1.3) o qual se encontra incorporado no sistema global.

3.1.1 Use cases

O modelo *use cases* (Fig. 3.1) é utilizado para representar as funcionalidades que a Duvideo deseja ver implementadas para os utilizadores interagirem com o sistema.

Sendo a Duvideo uma empresa de produção de conteúdos audiovisuais, as principais actividades que os seus utilizadores irão realizar serão relacionadas com:

- Introdução de novos vídeos no arquivo;
- Navegação pelo arquivo;
- Criação de metadados;
- Administração do sistema.

Neste sistema, existem diferentes utilizadores com diferentes privilégios de acesso a funcionalidades do mesmo. Os únicos actores que têm privilégios a nível de operações restritas são os Jornalistas, que têm a permissão para validarem os metadados que foram extraídos, quer proveniente das técnicas implementadas no protótipo, quer daqueles que são gerados automaticamente pela câmara. Os Documentalistas, que também são Jornalistas, têm ainda acesso a zonas de administração do sistema - e.g., actualizar o tesauro com novas categorias da normalização EUROVOC.

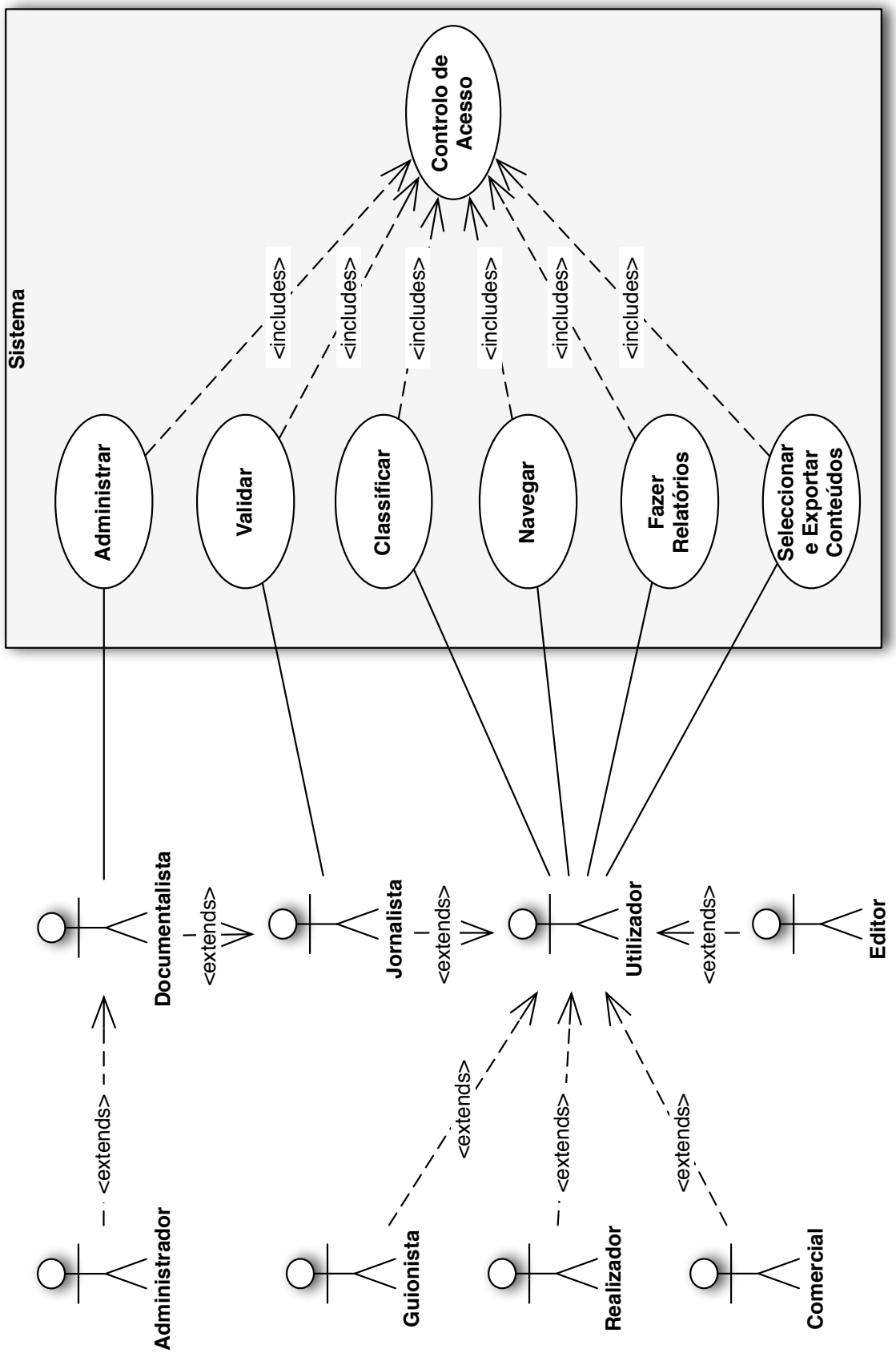


Figura 3.1: Modelo *Use Cases*.

3.1.2 Arquitectura do sistema proposta

Na figura 3.2 é possível visualizar a arquitectura proposta para o sistema, integrada com o *workflow* da Duvideo.

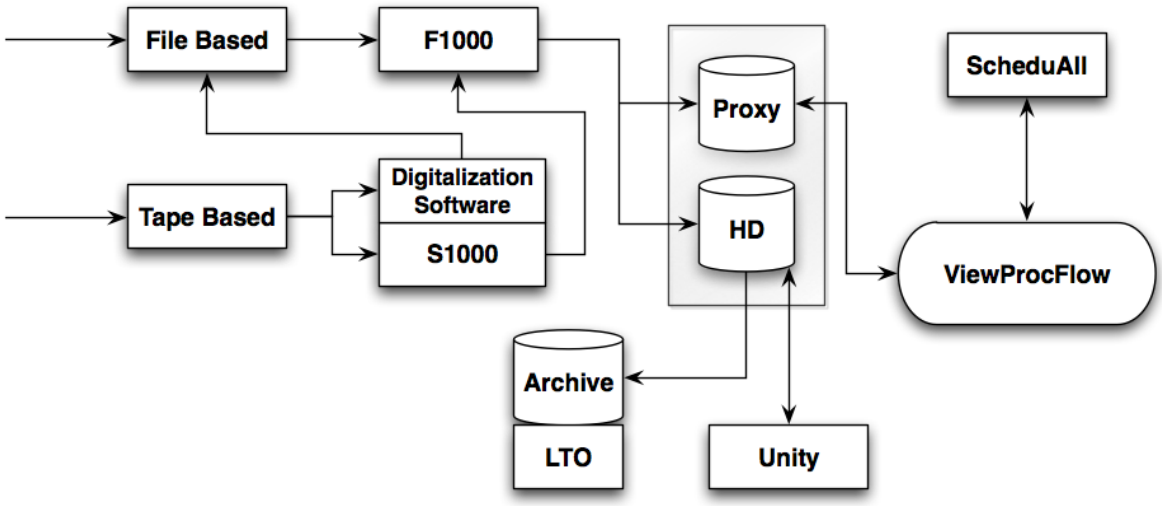


Figura 3.2: Localização do ViewProcFlow na arquitectura do sistema.

Os conteúdos que chegam ao sistema, podem vir num suporte analógico (que necessita uma conversão para digital) ou vindos já de num suporte digital. Para os vídeos ainda em suporte analógico, o componente S1000 mxfSPEEDRAIL é capaz de controlar Video Tape Recorder (VTR), via protocolo VDCP (Video Disk Control Protocol)[1], e fazer a conversão para várias resoluções. O F1000 mxfSPEEDRAIL é uma solução para mover os ficheiros MXF para AVID Unity ou MediaNetwork que suporta os principais dispositivos de gravação do mercado: Sony XDCAM, Panasonic P2TM, Thomson GV Inifity. Os componentes S1000 e F1000 integram uma interface SOAP (Simple Object Access Protocol) para fácil integração com outro sistemas. Este componente cria versões de qualidade mais reduzida dos vídeos originais, procedendo ao seu armazenamento nos servidores Proxy e HD. No servidor HD encontram-se as versões de alta qualidade que serão usadas na criação do produto final e no servidor Proxy encontram-se versões com uma qualidade inferior (consequentemente de menor dimensão) sobre as quais os utilizadores criam os guiões de edição[2].

[1]O protocolo VDCP, também conhecido por Louth Protocol, é usado principalmente para o controlo de servidores de video, onde o dispositivo de controlo tem a iniciativa da comunicação com o dispositivo controlado (disco de vídeo).

[2]Estes guiões serão utilizados, juntamente com as versões de alta qualidade, para criar o produto final.

Os metadados que forem gerados pelo ViewProcFlow, serão adicionados aos respectivos MXF de cada vídeo e ao ficheiro XML correspondente (secção 2.4.4).

O ScheduALL é uma ferramenta composta por vários componentes, entre os quais o Resource Management, que realiza a gestão de conteúdos e acesso ao Arquivo. Nele os utilizadores também podem introduzir manualmente informações, nos seguintes separadores:

Notes
Permite ao Documentalista criar notas sobre aspectos como qualidade, conservação, preservação e circunstâncias de utilização.

Cut/Details
Onde estão as descrições dos conteúdos.

Trafic History/Other
Uma área não interventiva, onde é registado o movimento de entrada e saída dos suportes do arquivo.

Schedulling Requests
Faz a ligação entre os vários departamentos, nomeadamente com a Produção.

Metadata
Reúne as modificações aos documentos.

Miscellaneous
Gestão de projectos.

3.1.3 Arquitetura do ViewProcFlow

O protótipo ViewProcFlow foi organizado num modelo cliente-servidor. No componente de Servidor, que irá receber para processar as versões Proxy dos vídeos, encontram-se todas as operações de extracção de metadados - i.e., segmentação, extracções de descritores, detecção de faces, detecção de conceitos - para serem utilizadas sobre os vídeos, assim como os serviços de comunicação com o componente Cliente. Do lado do Cliente encontram-se todas as funcionalidades de visualização do arquivo e metadados associados, assim como métodos para a validação dos últimos. Este será a interface para os utilizadores comunicarem com o sistema.

A figura 3.3, representa a arquitectura utilizada para o protótipo. Todas as comunicações de dados entre o Cliente e o Servidor são realizadas através do componente ofHttpServer, um *addon* à biblioteca openFrameworks [LWC09], com recurso a XML.

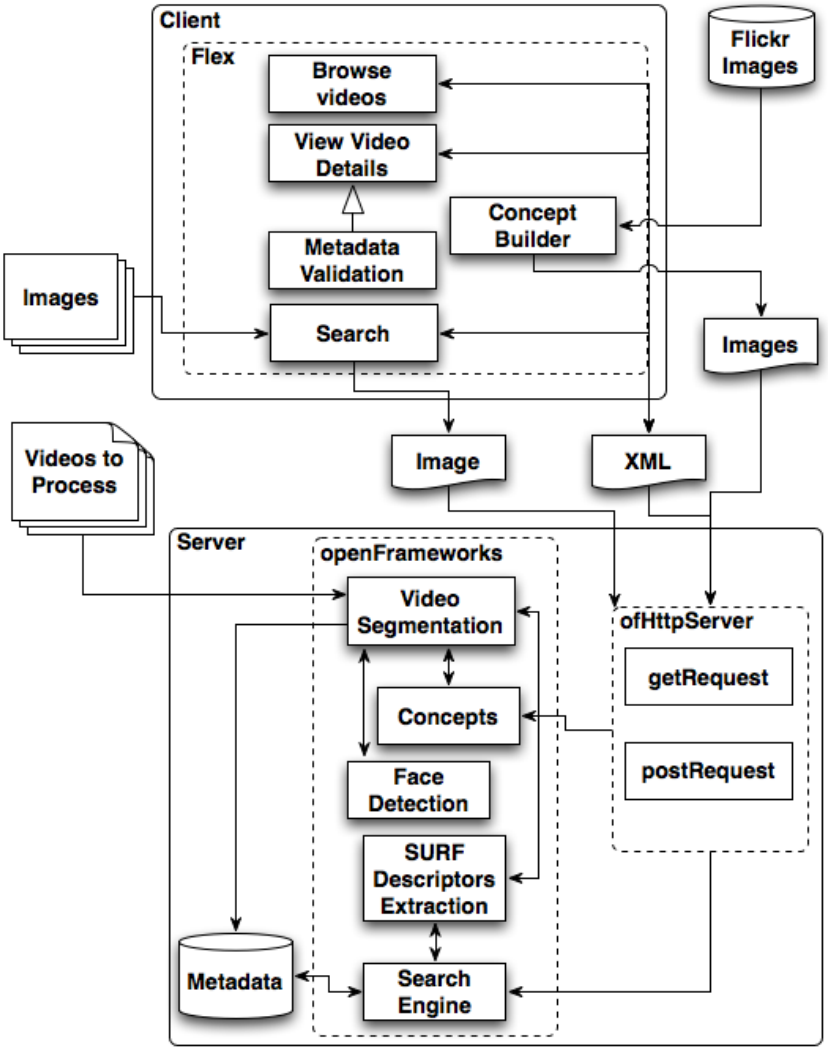


Figura 3.3: Arquitectura do ViewProcFlow.

Por exemplo, o pedido da listagem de todos os vídeos, inicia-se no Cliente através de uma chamada ao serviço "videosService" (ver listagem 3.1) e fica em espera da resposta do Servidor.

Listagem 3.1: Exemplo de código de um HTTPService em Flex 4.

```
<mx:HTTPService id="videosService"
            url="http://localhost:8888/getVideos.of"
            method="POST"
            result="resultHandler_VideosService(␣event␣)"
            fault="resultFaultHandler(␣event␣)"
            showBusyCursor="true"/>
```

Quando o servidor ofHttpServer recebe o pedido, processa-o e envia a resposta que neste caso será uma lista com os vídeos existentes (ver listagem 3.2).

Listagem 3.2: Exemplo de uma resposta do Servidor em formato XML.

```
<?xml version="1.0" encoding="UTF-8" ?>
<VIDEOS>
    <VIDEO>blackwhite.mov</VIDEO>
    <VIDEO>camel.mov</VIDEO>
    <VIDEO>concerto.mov</VIDEO>
    <VIDEO>ederly.mov</VIDEO>
    <VIDEO>landscape.mov</VIDEO>
    <VIDEO>landscape2.mov</VIDEO>
    <VIDEO>landscape3.mov</VIDEO>
    <VIDEO>little_girl.mov</VIDEO>
    ...
</VIDEOS>
```

No Cliente, a resposta será processada pela função "resultHandler_VideoService". No caso de existir algum erro na comunicação, a função "resultFaultHandler" recebe a exceção e apresenta o erro ao utilizador numa mensagem de alerta. Para a funcionalidade de criação de conceitos, a comunicação é idêntica, mudando o endereço do serviço. Além desta troca de informação em formato XML, passagem das imagens do Cliente para o Servidor em também é realizada através de um serviço onde a imagem é enviada como parâmetro (ver listagem 3.3).

Listagem 3.3: Exemplo do upload de uma imagem para o Servidor por ActionScript.

```
private function uploadFile( evt:Event ):void {
        try {
          var request:URLRequest =
                new URLRequest( http://localhost:8888/sendImage.of );
          request.method = URLRequestMethod.POST;
          var uploaderReqVars:URLVariables =
                new URLVariables( "var=samevalue" );
          request.data = uploaderReqVars;
          fileRef.upload( request, "file" );
        } catch ( err:Error ) {
          Alert.show( "Erro␣no␣envio␣do␣ficheiro",
              "Envio␣de␣ficheiro", Alert.OK"␣);}}
```

3.2 Realização

Nesta secção será descrito as funcionalidades e interfaces dos dois componentes da arquitectura, Servidor (secção 3.2.1) e Cliente (secção 3.2.2).

3.2.1 Servidor

Para desenvolvimento do protótipo da aplicação do servidor recorreu-se à biblioteca openFrameworks. As principais vantagens desta biblioteca são: ser uma API (Application Programming Interface) de fácil utilização para a manipulação de vídeo e imagem e conter uma camada abstracta que torna a aplicação independente do sistema de operação.

O servidor encontra-se em espera activa por novos vídeos que cheguem para serem processados. Quando se inicia o processamento de um novo vídeo, é gerado o "video.xml" (Fig. 3.4).

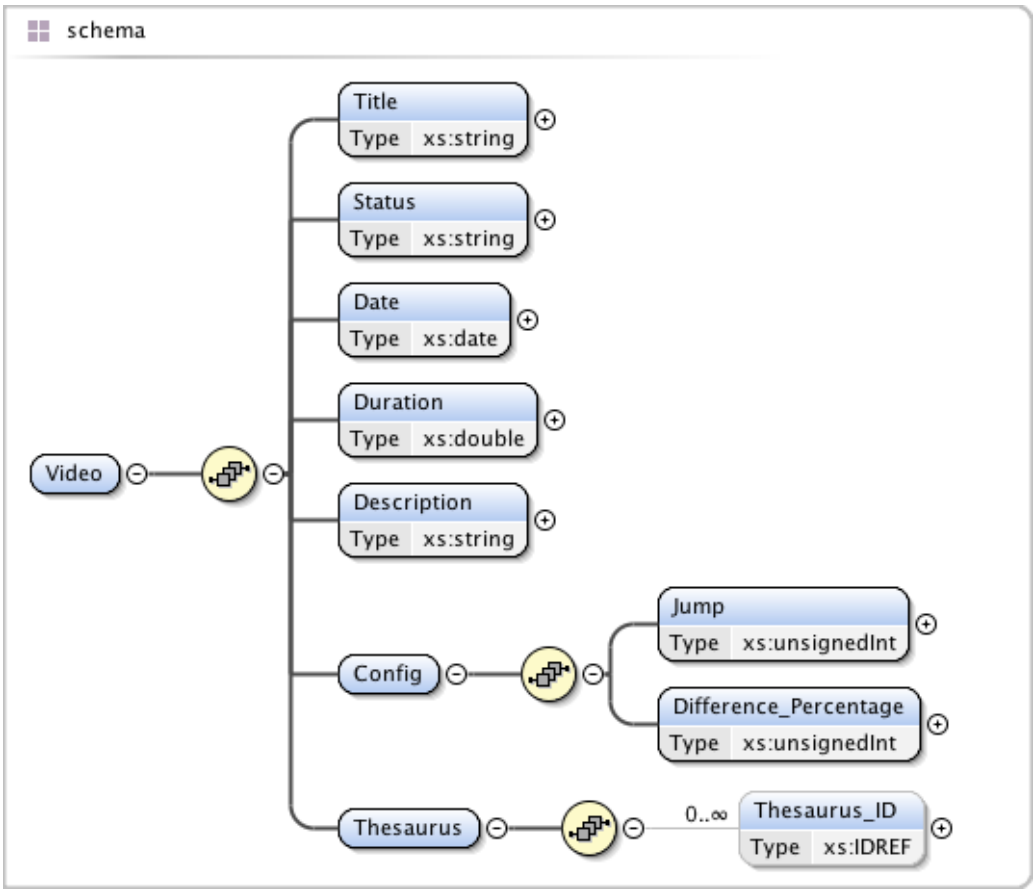


Figura 3.4: *Schema* para o XML "video.xml".

Este ficheiro XML vai conter além de campos tradicionais - e.g., *title*, *date*, *duration* e *description* - campos cuja a sua aplicação está relacionada com o domínio do problema.

O campo *Status* reflecte o actual estado do vídeo no que diz respeito à validação dos metadados, podendo conter os seguintes valores:

- **To Validate** - Uma vez processado o vídeo, este fica disponível para o utilizador validar os metadados extraídos. Neste caso, o utilizador que o irá validar é um jornalista.
- **Valid** - Caso o utilizador concorde com os dados extraídos, coloca o vídeo neste estado.
- **To Process** - Se os metadados contiverem muitos erros, do ponto de visto do utilizador, este pode pedir um novo processamento, alterando os parâmetros de configuração.

O nó *Config* contém os dados utilizados na segmentação. O campo *Difference_Percentage* influência o valor do *threshold*, já o *Jump* irá influenciar o número de frames a saltar após detectar uma mudança de cena. Por fim, o campo *Thesaurus* irá conter identificadores do EUROVOC - Thesaurus que não são possíveis de detectar individualmente em cada cena - i.e., só através de união do conceitos detectados em cenas distintas, será possível ter uma melhor compreensão do que esse conceito poderá representar. Depois da criação deste ficheiro XML, tem início a segmentação.

Segmentação

Com recurso ao algoritmo de diferença absoluta de histogramas (secção 2.2.1), os vídeos são processados e a informação dos segmentos detectados, é adicionada ao ficheiro "scene.xml" existente para cada vídeo. A sua estrutura está representada na figura 3.5.

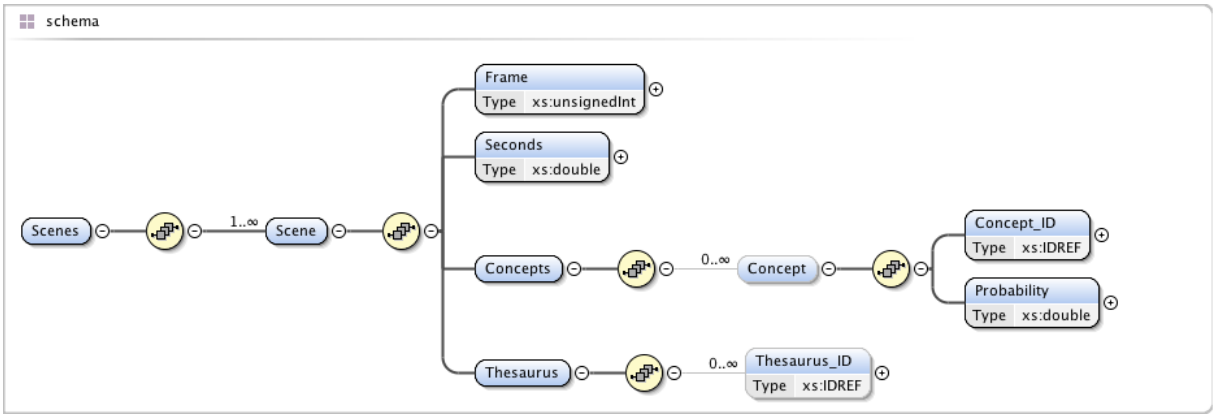


Figura 3.5: *Schema* para o XML "scenes.xml".

Visto as aplicações trabalharem com dados temporais em formato diferente, optou-se por armazenar a informação tanto em "segundos" como usando o "número da *frame*" em que a cena tem inicio.

A partir deste momento, temos as cenas detectadas e as imagens que as irão representar. É iniciado o processo de extracção de metadados para cada uma dessas imagens: descritores, faces e conceitos.

Extracção de descritores

São extraídos os descritores gerados pelo método SURF (secção 2.3.2) com recurso aos métodos do OpenCV [ope10], existentes no openFrameworks, para este tipo de descritores. A informação é armazenada no ficheiro XML (Fig. 3.6) respeitante à imagem representativa da cena - i.e., para a cena "2" será criado o ficheiro "surf-s2.xml".

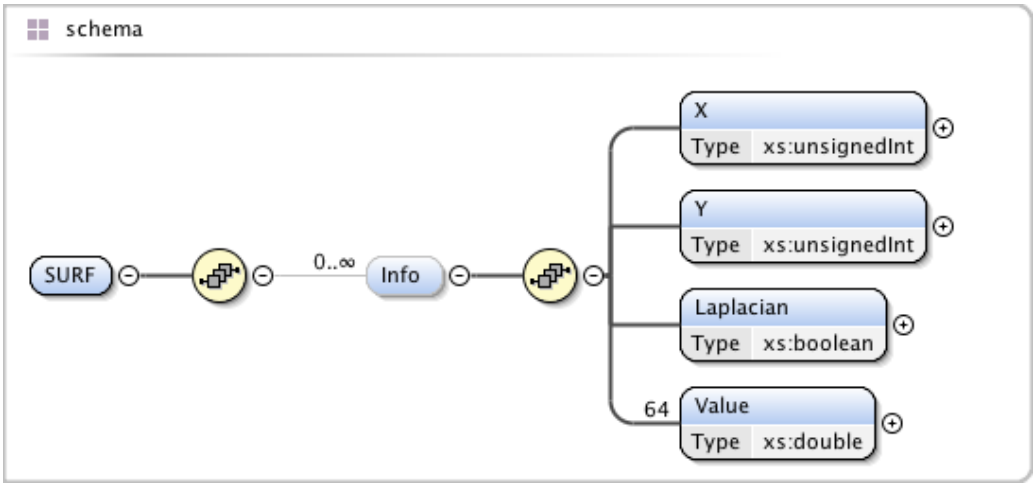


Figura 3.6: *Schema* para o XML "surf-s*i*.xml".

Detecção de faces

Esta detecção é realizada através da técnica designada por Viola Jones (secção 2.3.3) com o código disponibilizado pelo grupo de investigação Lalalab [BD10]. Todos os dados relevantes de cada face que é encontrada, são armazenados no ficheiro "faces.xml" respectivo ao vídeo (Fig. 3.7). Adicionalmente é extraída também a imagem da face, para melhorar a utilização final da interface. Como podem existir várias faces na mesma cena, o campo *Index* fará a distinção entre estas. Visto a detecção do género da face ainda não se encontrar implementado, este campo - *Gender* - tem o valor por omissão de indefinido.

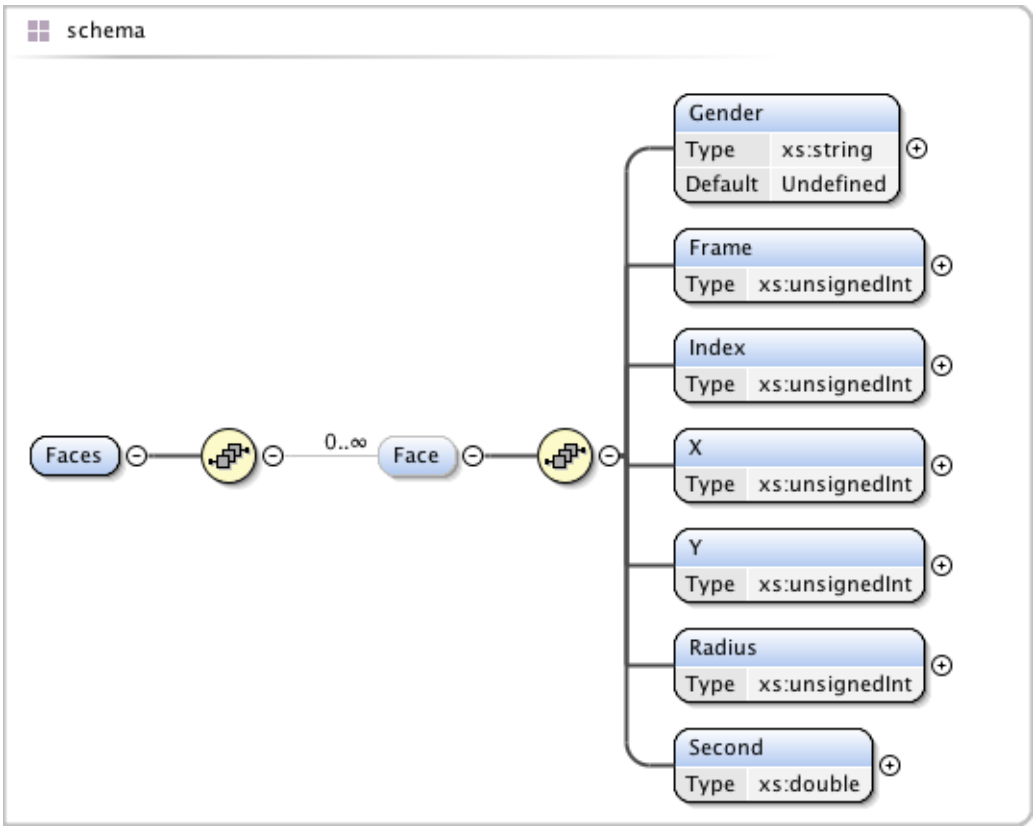


Figura 3.7: *Schema* para o XML "faces.xml".

Detecção de conceitos

Com recurso ao método descrito na secção 2.3.4.1, cada imagem é avaliada pelos classificadores de forma a detectar a presença dos conceitos previamente treinados. Sempre um conceito numa imagem representativa de uma cena for detectado, este é adicionado à estrutura do ficheiro XML das cenas, no nó "Concept" (Fig. 3.5), assim como a sua probabilidade de ocorrência indicada pelo classificador.

3.2.2 Cliente

O Cliente está a ser desenvolvido em Flex 4.0. Com uma base em MXML, uma extensão de XML, para a construção das interfaces e com recurso ao ActionScript para as componentes lógicas, esta tecnologia proporciona uma fácil abordagem de prototipagem rápida para a interface, sem por em risco as funcionalidades esperadas. A separação do conteúdo da visualização foi um também aspecto levado em conta desde o início, por isso todo o conteúdo encontra-se alugado no Servidor e é pedido no inicio da execução através de serviços ligados ao ofHttpServer - e.g., o catálogo do tesauro.

De modo a tirar partido dos metadados produzidos no Servidor, a interface de visualização é um factor chave para os utilizadores. As especificações preliminares foram baseadas nas descrições iniciais dos utilizadores. Assim sendo, a interface encontra-se dividida em:

- Navegação e pesquisa no arquivo de vídeo
- Visualização de vídeo e metadados
- Gestão de ontologias e conceitos

A janela inicial mostra o arquivo de vídeo do lado direito com os parâmetros de pesquisa à esquerda (Fig. 3.8). As opções de pesquisa disponíveis ao utilizador, são as

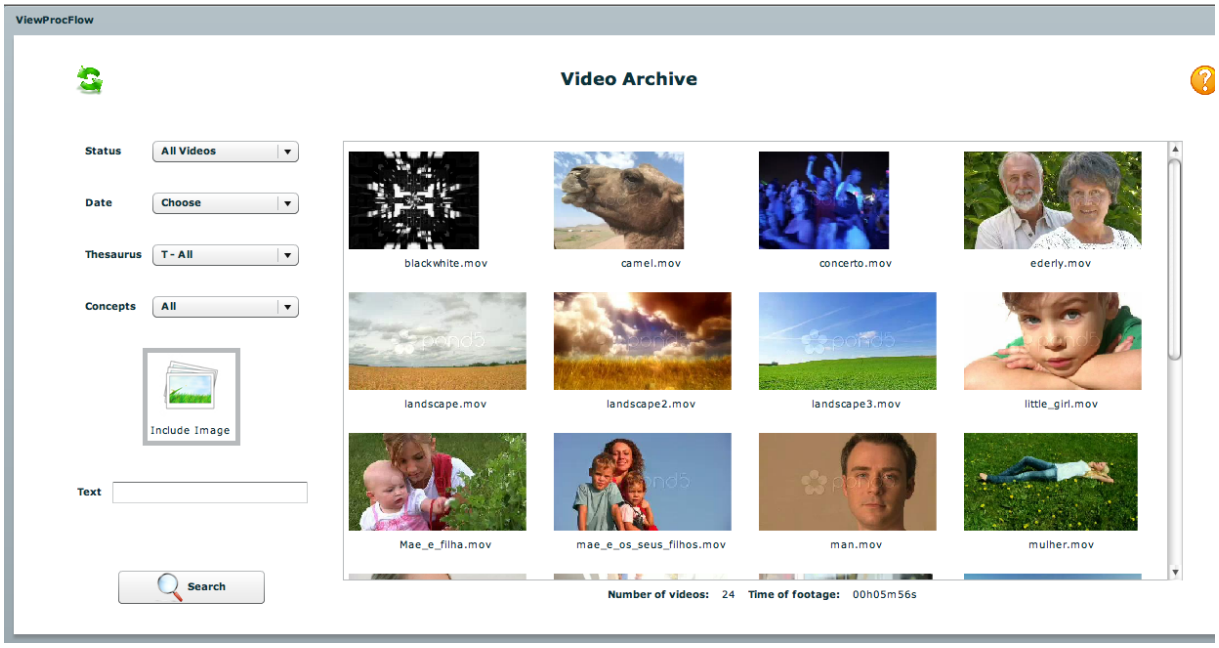


Figura 3.8: Janela principal do ViewProcFlow.

seguintes: *a*) *Status*, permitindo escolher vídeos já validados, não validados ou todos os vídeos *b*) *Date*, o utilizador tem a hipótese de escolher um intervalo de datas; videos anteriores ou posteriores a uma data *c*) *Thesaurus*, permite a escolha das categorias do EUROVOC *d*) *Concepts*, o utilizador pode escolher quais conceitos que está interessado *e*) *Image*, o utilizador pode escolher uma imagem que deseja procurar no arquivo de vídeo *f*) *Text Input*, pesquisa textual em todos os campos, desde o título às anotações .

Quando no campo de pesquisa o utilizador deseja procurar por uma imagem específica, pode escolher uma imagem já armazenada no sistema ou então fazer o *upload* de uma nova imagem. Ainda assim é possível que essa imagem contenha elementos

que não são o objecto de pesquisa e dessa forma é fornecido um editor de imagem (Fig. 3.9). O utilizador pode seleccionar qual a área da imagem pela qual realmente

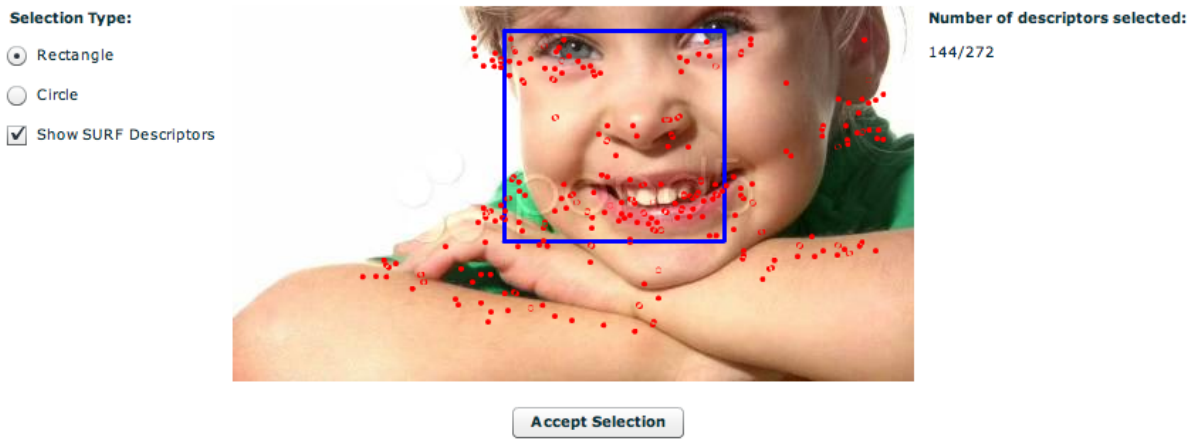


Figura 3.9: Editor de Imagem.

deseja pesquisar e ainda visualizar os pontos que serão utilizados na pesquisa. Desta forma, o utilizador tem a informação necessária para não escolher zonas que tenham poucos pontos de interesse. De modo a que os resultados destas pesquisas baseadas em imagem sejam mais flexíveis, o utilizador tem um parâmetro para indicar a percentagem de pontos necessários para que uma cena seja considerada como resultado final. Assim, com o exemplo da figura 3.9, se o utilizador escolher uma percentagem muito elevada o resultado será muito semelhante à imagem base. Caso esta percentagem seja mais baixa, mais resultados podem aparecer - e.g., uma cena onde a criança continua a sorrir mas está de olhos fechados 3.10.

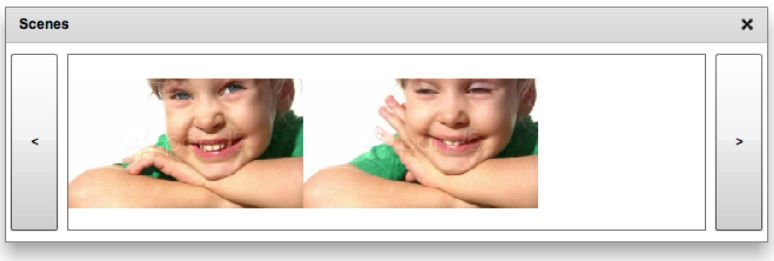


Figura 3.10: Exemplo de um resultado de uma pesquisa com base numa imagem.

Do lado direito, são dispostos os vídeos existentes no arquivo, ou os resultados das pesquisas efectuadas. Caso esses vídeos apresentados sejam relativos a uma pesquisa, quando é seleccionado um dos vídeos, pertencentes ao resultado, é lançado um *popup*

(Fig. 3.10) com a lista de cenas do vídeo que foram marcadas como contendo dados pesquisados.

Quando se escolhe um dos vídeos apresentados, passamos para um ambiente em que se podem observar os metadados do vídeo (Fig. 3.11). Neste ambiente os meta-

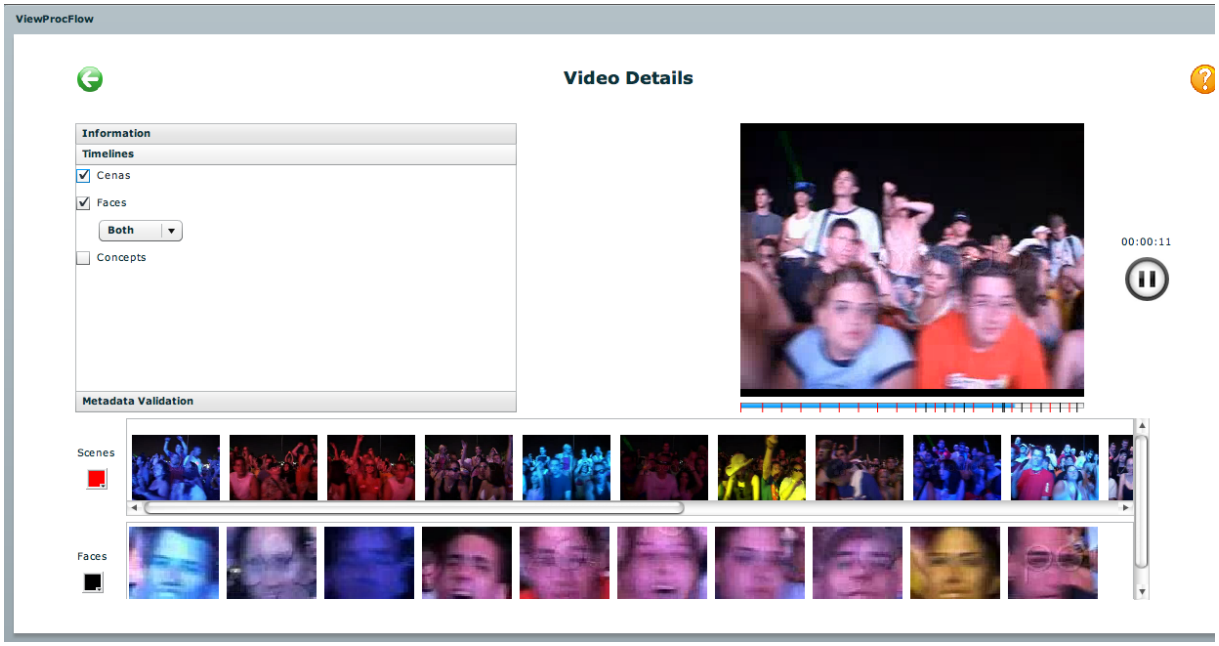


Figura 3.11: Ambiente de visualização do vídeo e metadados extraídos.

dados - e.g., cenas, faces, conceitos - são representados em *timelines*. Cada um destes elementos funciona como âncoras para o local temporal onde ocorre no vídeo, de modo a facilitar a visualização do mesmo. Quando o utilizador seleccionar uma dessas âncoras, o vídeo é colocado na posição temporal onde a mesma ocorre. Uma segunda forma de observar a localização onde ocorrem estes elementos é na barra de progresso onde são colocados marcadores da cor do tipo da *timeline* escolhida.

Após a visualização dos metadados, o utilizador pode submeter a sua validação ou então pedir para que o vídeo volte a ser processado modificando os parâmetros do algoritmo (Fig. 3.4). Para casos em que são poucos os erros encontrados nos metadados, o utilizador pode remover o elemento através do seu menu de contexto.

Relativamente à criação de novos conceitos, o ambiente foi dividido em duas partes. A componente referente à visualização da ontologia incoorporada no EUROVOC pode ser observada na figura 3.12. Com recurso à biblioteca SpringGraph [She10], a ontologia é disposta num grafo. O utilizador tem a possibilidade de navegar, pesquisar por nós e aumentar a semântica acrescentando nós ao grafo que representa a ontologia.

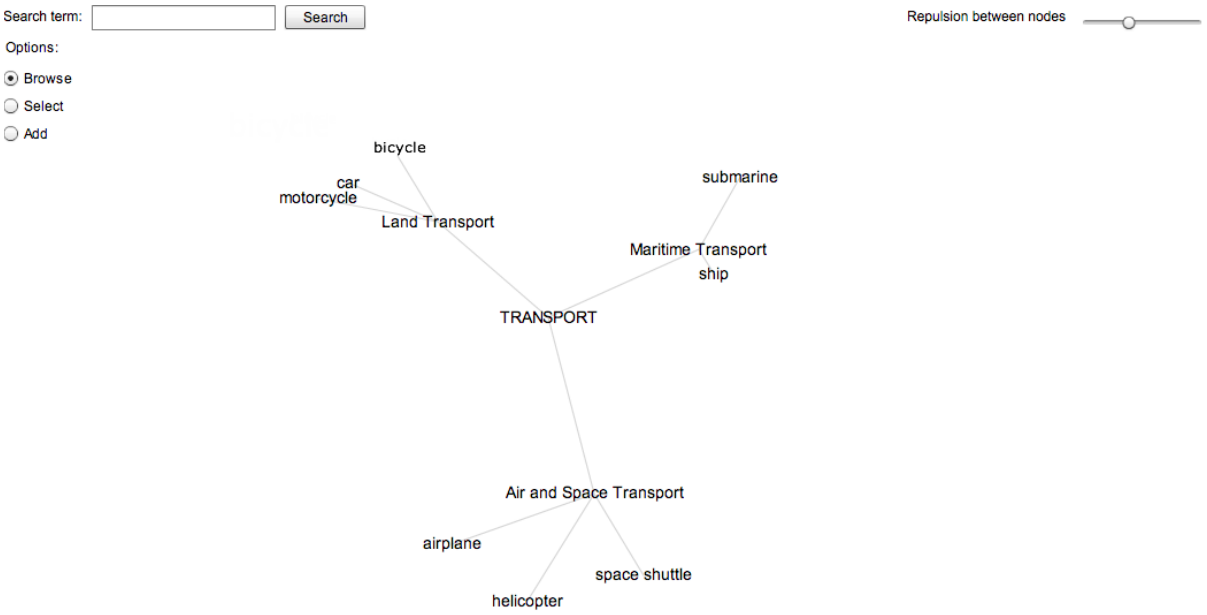


Figura 3.12: Ambiente de visualização do EUROVOC e ontologia.

Como foi descrito no capítulo anterior, a ontologia pode ajudar o utilizador a criar os conjuntos de testes, sugerindo termos para o conjunto de negativo. A primeira abordagem para as sugestões apresentadas pela ontologia, são termos que se encontram ao mesmo nível e ligados ao mesmo nó - e.g., para o conceito "car" são sugeridos "bicycle" e "motorcycle", que são os termos relacionados com o nó "Land Transport".

O ambiente para a criação de um novo conceito (Fig. 3.13) contém as sugestões da ontologia com base no conceito que vamos treinar. Com recurso a uma ligação ao Flickr [fli10], um repositório de imagens, o utilizador pode pesquisar por imagens e arrastá-las para o conjunto positivo ou negativo respectivamente. Depois de atribuir um nome ao conceito, os conjuntos de imagens são enviados para o servidor, de modo a treinar o novo classificador.

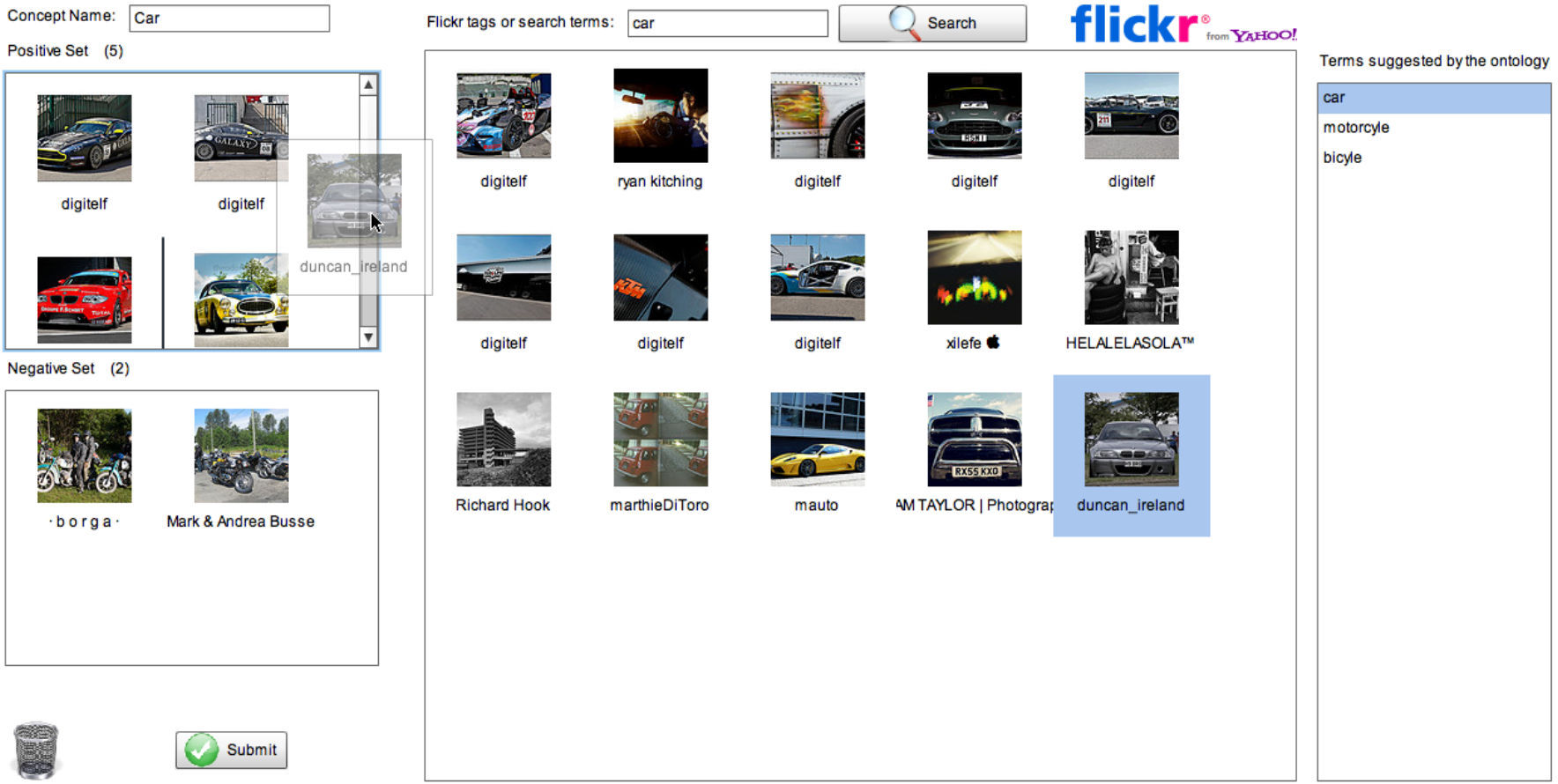


Figura 3.13: Ambiente para construir novos conceitos

3.3 Avaliação

Esta secção contém os testes realizados ao protótipo, tanto a nível de desempenho como de resultados. São também apresentados os resultados dos inquéritos realizados aos utilizadores que conduziram alguns testes ao protótipo.

3.3.1 Resultados das técnicas aplicadas

Cada uma das funcionalidades de extracção de metadados foi avaliada de forma a validar os resultados apresentados.

Segmentação

A segmentação é uma funcionalidade onde existe uma disparidade nos resultados finais, devido à variedade de conteúdo possível num vídeo - i.e., o resultado de cenas de um vídeo com muito movimento é diferente de um vídeo com pouco movimento. No entanto, existe a possibilidade de o utilizador modificar os parâmetros da segmentação e através de poucas iterações conseguir chegar a bons resultados.

Detecção de faces

O algoritmo utilizado produz excelentes resultados finais, com um nível bastante reduzido de falsos positivos. Devido ao classificador ter sido treinado para faces completas e de uma perspectiva relativamente frontal, fará com que haja alguns falsos negativos nomeadamente em faces de perfil.

Extracção de descritores

A utilização dos descritores SURF revelou excelentes resultados na pesquisa. Para graus de precisão muito altos (opção parametrizada pelo utilizador) e com poucas dezenas de descritores o resultado é exacto.

Extracção de conceitos

Para extracção dos conceitos “Face”, “Pessoas”, “Praia”, “Espaços Interiores”, “Festa”, “Neve” e “Natureza” foram alcançados com bons resultados. Já foi dado inicio ao processo para a criação dos novos conceitos que foram enunciados na tabela 2.1, através da interface (Fig. 3.13) e ontologias (Fig. 3.12) descritas anteriormente.

3.3.2 Testes de desempenho do protótipo

Os testes foram conduzidos num Macbook com as seguintes características:

- Processador Intel Core 2 Duo 2,4GHz
- Memória de 4GB DDR3 a 1066MHz
- Disco rígido a 5400rpm

Para realizar os testes de extracção de metadados foram utilizados vídeos de pequena duração, com 40 segundos a 4 minutos e com dimensões de 320x240 a 640x320. A tabela 3.1 reflete os tempos de execução das principais tarefas.

Tabela 3.1: Tarefas e os seus tempos de execução.

Tarefa	**Média de Tempos de Execução**
Diferença de Histograma Entre Duas Imagems	0.17s
Detecção de Faces	0.02s
Extracção de Descritores SURF	0.94s
Comparação de Descritores Entre Duas Imagens	0.21s
Detecção de Conceitos numa Imagem	25s

Em todo este processo, a segmentação introduz bastantes benefícios, visto reduzir a redundância de informação de onde se irão extrair metadados.

A pesquisa por imagem é uma funcionalidade que utiliza os descritores como factor de comparação. Mediante a percentagem de precisão que o utilizador indicar e sem quaisquer outros parâmetros disponíveis nas pesquisas, o tempo de resposta de uma pesquisa pode ser da ordem de minutos para uma biblioteca de pequenas dimensões.

Relativamente à detecção de conceitos, o tempo de execução é mais considerável que as outras operações, no entanto isto deve-se a uma API inicial para testes, onde alguns aspectos de desempenho não eram considerados. Toda a informação calculada para a imagem não estava a ser armazenada, por isso a cada nova iteração tinha que ser calculada. Outro aspecto que aumenta o tempo de execução está relacionado com o teste de todos os sete conceitos (“Face”, “Pessoas”, “Praia”, “Espaços Interiores”, “Festa”, “Neve” e “Natureza”) não havendo uma árvore de decisão para quais os conceitos a detectar mediante o resultado positivo de algum conceito - e.g., dada uma imagem ter uma forte possibilidade de ocorrência do conceito de “Espaço Interior”, o conceito “Natureza” terá uma menor correlação daí não ser um conceito a testar no imediato.

3.3.3 Inquérito

Foi realizado um inquérito (ver anexo C) composto por sete tarefas, com o objectivo de validar as funcionalidades implementadas assim como a usabilidade da interface e sete questões de avaliação das funcionalidades, a nível de desempenho e resultados e também das suas interfaces. Os oito utilizadores seleccionados para este inquérito tinham experiência tanto em tecnologias de informação ligadas à multimédia como alguns conhecimentos de manipulação de conteúdos audiovisuais. Antes de ser efectuado o inquérito ao utilizador, foi dada uma introdução à aplicação, explicando as suas funcionalidades e objectivos pretendidos.

Relativamente às tarefas pedidas, os utilizadores não tiveram dificuldades em as executar e avaliaram os resultados obtidos com os esperados de forma positiva. Houve uma tarefa, no entanto, em que foram levantadas questões esperadas ao nível da subjectividade dos resultados. Na tarefa 2, é pedido ao utilizador para verificar os metadados extraídos de um vídeo e pedir um novo processamento dos mesmo. O conteúdo deste vídeo é um grande plano de uma criança, em que esta muda de expressão várias vezes ao longo do vídeo. Alguns utilizadores entenderam que tudo era a mesma cena e concordaram com uma nova segmentação do vídeo mas outros aceitaram a segmentação e por sua vez os metadados daí extraídos, pois no seu entender a mudança de expressão da criança deveria indicar uma nova cena. Este exemplo foi utilizado para mostrar a subjectividade natural do observador na avaliação do conteúdo de um vídeo.

A tarefa 6 foi alvo de algumas críticas de usabilidade. Apesar de a opinião geral ser positiva em relação à disposição da ontologia num grafo, a forma de interagir com este não foi consensual. Foi sugerido que o método de interacção com o grafo, através dos *radio buttons*, fosse alterado para um menu de contexto para tornar sua utilização mais fluida. Outro aspecto indicado, foi a falta de ligação com os conceitos já existentes - i.e., o utilizador devia ter a informação de quais os termos na ontologia que já incluem classificadores para detectar esse conceito.

Nas avaliações pedidas, tanto as funcionalidades como as interfaces tiveram avaliações positivas (ver anexo D). O único caso onde existiu uma avaliação negativa, foi em relação ao desempenho da pesquisa com imagem, devido a ter um tempo de resposta muito grande, apesar da avaliação dos resultados desta pesquisa serem positivos.

Existiram sugestões interessantes por parte dos inquiridos ao nível de funcionalidades, interfaces e tecnologias. Devido ao processo de extracção de metadados ser demorado, foi sugerido que caso o utilizador peça um novo processamento ao vídeo, os antigos metadados sejam mantidos até uma aceitação dos novos. Caso os novos resultados tenham uma qualidade inferior, seria possível retroceder aos antigos sem a

necessidade de fazer um novo processamento. Um pedido, que já fazia parte da lista de tarefas para a próxima versão do protótipo, seria um editor de vídeo e a sugestão do YouTube Editor como uma interface possível (Fig. 3.14).

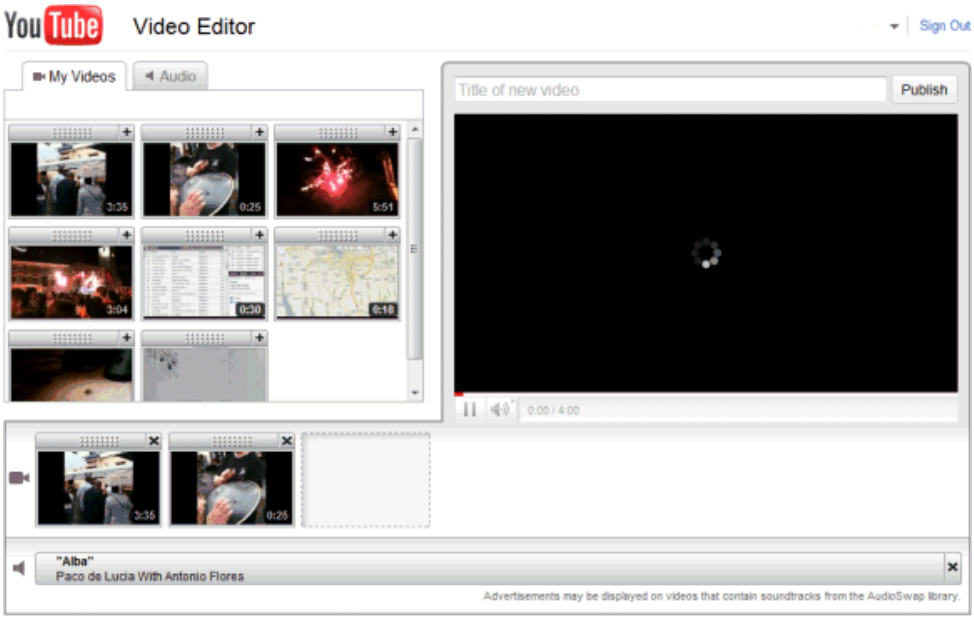


Figura 3.14: Interface do YouTube Editor.

Relativamente às tecnologias, houve a sugestão da possibilidade de utilização do HTML5 para o desenvolvimento do Cliente, visto estar bem preparado para o suporte de vídeo e imagem e ter desempenho melhor que o Flash. Também foi sugerido um estudo à biblioteca Open Source Media Framework [osm10], que permite a construção de leitores de vídeo, e em que a incorporação de metadados se encontra facilitada, o que poderia despoletar novas funcionalidades - e.g., hiperligações no próprio vídeo para outros conteúdos.

4

Conclusões e trabalho futuro

Este capítulo descreve as conclusões retiradas do trabalho efectuado e os pontos fundamentais a ter em conta para o trabalho futuro no desenrolar do projecto VideoFlow.

4.1 Conclusões

Foi realizado um sistema para facilitar a produção de conteúdos audiovisuais, tornando os processos de criação mais eficientes. Para tal, a primeira versão proposta envolve a integração de diferentes tecnologias de extracção de metadados de modo a serem integradas no *workflow* de produção de vídeo.

O processo de detecção de cenas de um vídeo é uma funcionalidade muito importante devido à enorme quantidade de material vídeo não segmentado em formato analógico. Apesar deste processo de segmentação ter algumas falhas e requerer alguma correcção humana, já introduz uma enorme poupança a nível de recursos humanos e de tempo dispendido. Relativamente à extracção de descritores, a sua utilização nas pesquisas com base em imagens revelou bons resultados, apesar do tempo necessário para os atingir. Por fim, a detecção de conceitos integrada com o tesauro mostrou ajudar a anotação de conteúdos, ainda que a construção da biblioteca de classificadores esteja no início.

Para os testes realizados, os vídeos encontravam-se no formato MOV (Quicktime), que é um formato diferente do esperado MXF. É necessário a utilização de uma outra

biblioteca para a manipulação deste tipo de ficheiros. A solução criada pela MOG Solutions [MOG09] poderá ser uma das possibilidades, no entanto não houve a possibilidade de ter acesso à mesma durante a realização da dissertação. Foi também realizada uma primeira integração da biblioteca disponibilizada pelo FreeMXF [fre09] mas sem sucesso devido ao código não estar preparado para os sistemas operativos actuais.

Foi avaliado, ainda de que uma forma preliminar, o protótipo por um conjunto de utilizadores com bons resultados e que deixaram sugestões de funcionalidades para próximas versões do protótipo.

Apesar deste estudo ainda se encontrar no início, é possível verificar que as tecnologias aqui estudadas, podem enriquecer a semântica de dados extraídos dos arquivos de vídeo, com processo automáticos ou semi-automáticos e por sua vez melhorar o *workflow* onde são incluídas.

4.2 Trabalho futuro

Como trabalho futuro, a integração completa do protótipo com o formato MXF é um aspecto importante para o desenrolar do projecto VideoFlow.

Ao nível das técnicas utilizadas para a extracção de metadados, a área da segmentação de vídeo está em constante evolução e novos algoritmos ou modificações a antigos, são apresentados constantemente por isso é uma área a considerar no futuro. Relativamente aos conceitos semânticos, a continuação do trabalho já realizado na construção da ontologia e regras que possam tirar partido dos conceitos de forma a enriquecer os dados é um trabalho muito importante. Como já foi dito anteriormente, é difícil descrever conceitos abstractos ou ambíguos e através destas regras será possível avançar num caminho onde mais conceitos possam ser detectados. Em paralelo, deve prosseguir o treino de mais classificadores para conceitos, de forma a aumentar a biblioteca de classificadores disponíveis no sistema e aumentar assim as possibilidades de anotações.

Nas tarefas existentes, todo o processamento é realizado sequencialmente, o que com a introdução de computação paralela, traria aumento do desempenho com um benefício de relevo nos tempos de execução [SSK+05]. De uma maneira similar, a utilização de uma base de dados nativa em XML - e.g., sedna [MOD10] - melhoraria o acesso aos dados, assim como as pesquisas possíveis nestes mesmos dados através de XPath e XQuery.

O reconhecimento de faces é uma funcionalidade interessante que irá enriquecer o protótipo, pois além de detectar a presença de pessoas, a possibilidade de as identificar fará com que as pesquisas possam ser melhor direccionadas.

A criação de histórias, através da edição dos conteúdos será também uma das funcionalidades a ser incorporada na próxima versão do protótipo.

Estas serão as linhas que irão guiar o trabalho futuro deste protótipo, de forma a torná-lo numa ferramenta mais robusta para colmatar as necessidades de um *workflow* de produção de conteúdos audiovisuais.

Bibliografia

[BD10] Clara Boj e Diego Díaz. lalalab - AR Magic System. http://www.lalalab.org/armagic.htm, Julho 2010.

[BTG06] Herbert Bay, Tinne Tuytelaars, e Luc J. Van Gool. Surf: Speeded up robust features. In *ECCV (1)*, pág. 404–417, 2006.

[Cab06] Susana Cabaço. VideoZapper - Um Sistema para Criação de Conteúdo de Vídeo Personalizado. Tese de Mestrado, Universidade Nova de Lisboa - Faculdade Ciências e Tecnologia, 2006.

[CGP+00] Patrick Chiu, Andreas Girgensohn, Wolf Polak, Eleanor Rieffel, e Lynn Wilcox. A genetic algorithm for video segmentation and summarization. In *In IEEE International Conference on Multimedia and Expo*, pág. 1329–1332, 2000.

[CLL08] Liang-Hua Chen, Yu-Chun Lai, e Hong-Yuan Mark Liao. Movie scene segmentation using background information. *Pattern Recognition*, 41(3):1056 – 1065, 2008. Part Special issue: Feature Generation and Machine Learning for Robust Multimodal Biometrics.

[cyc10] Cyc Knowledge Base. http://www.cyc.com/, Junho 2010.

[DAE95] Apostolos Dailianas, Robert B. Allen, e Paul England. Comparison of automatic video segmentation algorithms. In *In SPIE Photonics West*, pág. 2–16, 1995.

[Dev02] Bruce Devlin. The Material eXchange Format. 2002.

[DWBT06] Bruce Devlin, Jim Wilkinson, Matt Beard, e Phil Tudor. *The MXF Book: Introduction to the Material eXchange Format*. Elsevier, Março 2006.

[eur10a] EUROVOC - Thesaurus. http://europa.eu/eurovoc/, Junho 2010.

[Eur10b] União Europeia. Serviço de publicações. http://publications.europa.eu/, Junho 2010.

[fli10] Flickr. http://www.flickr.com/, Junho 2010.

[fre09] FreeMXF. http://freemxf.org/, Dezembro 2009.

[GJC09] Filipe Grangeiro, Rui Jesus, e Nuno Correia. Face recognition and gender classification in personal memories. In *ICASSP '09: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pág. 1945–1948, Washington, DC, USA, 2009. IEEE Computer Society.

[GPF98] Y. Gong, G. Proietti, e C. Faloutsos. Image indexing and retrieval based on human perceptual color clustering. In *CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pág. 578, Washington, DC, USA, 1998. IEEE Computer Society.

[HLMS95] Rune Hjelsvold, Stein Langørgen, Roger Midtstraum, e Olav Sandstå. Integrated video archive tools. In *IN PROCEEDINGS OF ACM MULTIMEDIA '95*, pág. 283–293, 1995.

[ima10] ImageCLEF - Image Retrieval in CLEF . http://www.imageclef.org/2010, Junho 2010.

[Jes09] Rui Jesus. *Recuperação de Informação Multimédia em Memórias Pessoais*. Tese de Doutoramento, Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologias, Setembro 2009.

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[LWC09] Zach Lieberman, Theodore Watson, e Arturo Castro. openFrameworks. http://www.openframeworks.cc/, Outubro 2009.

[MOD10] MODIS. sedna - Native XML Database System. http://modis.ispras.ru/sedna/, May 2010.

[mog05] XML Schema for MXF Metadata. Relatório técnico, MOG Solutions, Fevereiro 2005.

[MOG09] MOG - Solutions. http://www.mog-solutions.com, Dezembro 2009.

[NCHS10] Milind Naphade, Shih-Fu Chang, Alex Hauptmann, e John R. Smith. Large Scale Concept Ontology For Multimedia. http://www.lscom.org/, Junho 2010.

[NST+06] M. Naphade, J.R. Smith, J. Tesic, Shih-Fu Chang, W. Hsu, L. Kennedy, A. Hauptmann, e J. Curtis. Large-scale concept ontology for multimedia. *Multimedia, IEEE*, 13(3):86 –91, july-sept. 2006.

[ope10] openCV. http://opencv.willowgarage.com/wiki/, Julho 2010.

[osm10] Open Source Media Framework. http://www.opensourcemediaframework.com, Julho 2010.

[San06] Ernesto Santos. AAF, MXF, XML... Putting it all together. International Broadcast Conference, 2006.

[She10] Mark Shepherd. SpringGraph. http://mark-shepherd.com/SpringGraph/, Maio 2010.

[SMP09] The Society of Motion Picture and Television Engineers. http://www.smpte.org/, Dezembro 2009.

[SS02] Phillipe Salembier e Thomas Sikora. *Introduction to MPEG-7: Multimedia Content Description Interface*. John Wiley & Sons, Inc., New York, NY, USA, 2002.

[SS10] Cees G. M. Snoek e Arnold W. M. Smeulders. Visual-concept search solved? *IEEE Computer*, 43(6):76–78, June 2010.

[SSK+05] F. J. Seinstra, C. G. M. Snoek, D. Koelma, J. M. Geusebroek, e M. Worring. User transparent parallel processing of the 2004. In *NIST TRECVID Data Set," Proc. Int'l Parallel Distribution Processing Symp., 2005. ET AL.: THE SEMANTIC PATHFINDER: USING AN AUTHORING METAPHOR FOR GENERIC MULTIMEDIA INDEXING 1689*, 2005.

[SSW10] Arnold Smeulders, Cees Snoek, e Marcel Worring. MedialMill - Semantic Video Search Engine. http://www.science.uva.nl/research/mediamill/, July 2010.

[SWG+04] Cees G. M. Snoek, Marcel Worring, Jan-Mark Geusebroek, Dennis C. Koelma, e Frank J. Seinstra. The MediaMill TRECVID 2004 semantic video search engine. In *Proceedings of the 2nd TRECVID Workshop*, Gaithersburg, USA, November 2004.

[tre10] TREC Video Retrieval Evalution. http://trecvid.nist.gov/, June 2010.

[VJ01] Paul Viola e Michael Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.

[WC02] Howard D. Wactlar e Michael G. Christel. Digital video archives: Managing through metadata. pág. 80–95, Abril 2002.

[WCGH99] Howard D. Wactlar, Michael G. Christel, Yihong Gong, e Alexander G. Hauptmann. Lessons learned from building a terabyte digital video library. *Computer*, 32(2):66–73, 1999.

[Wen99] Robin Wendler. Library Digital Initiative Update: Metadata in the Library, Julho/Agosto 1999.

[wor10] WordNet - A Lexical database for English. http://wordnet.princeton.edu/, July 2010.

[YYL98] Minerva Yeung, Boon-Lock Yeo, e Bede Liu. Segmentation of video by clusthering and graph analysis. *Computer Vision and Image Understanding*, 71(1):94–109, July 1998.

EUROVOC - Thesaurus

04 ACTIVIDADE POLÍTICA

0406 quadro político
0411 partido político
0416 processo eleitoral
0421 assembleia
0426 trabalhos parlamentares
0431 vida política e segurança pública
0436 poder executivo e administração pública

08 RELAÇÕES INTERNACIONAIS

0806 política internacional
0811 política de cooperação
0816 equilíbrio internacional
0821 defesa

10 COMUNIDADES EUROPEIAS

1006 instituições da União Europeia e função pública europeia
1011 direito da União Europeia
1016 construção europeia
1021 finanças comunitárias

12 DIREITO

1206 fontes e ramos do direito
1211 direito civil
1216 direito penal
1221 justiça
1226 organização da justiça
1231 direito internacional
1236 direitos e liberdades

16 ACTIVIDADE ECONÓMICA

1606 política económica
1611 crescimento económico
1616 regiões e política regional
1621 estrutura económica
1626 contabilidade nacional
1631 análise económica

20 INTERCÂMBIOS ECONÓMICOS E COMERCIAIS

2006 política comercial
2011 política aduaneira
2016 trocas comerciais
2021 comércio internacional
2026 consumo
2031 comercialização
2036 distribuição comercial

24 FINANÇAS

2406 relações monetárias
2411 economia monetária
2416 instituições financeiras e crédito
2421 livre circulação de capitais
2426 financiamento e investimento
2431 seguros
2436 finanças públicas e política orçamental
2441 orçamento
2446 fiscalidade
2451 preços

28 QUESTÕES SOCIAIS

2806 família
2811 migrações
2816 demografia e população
2821 quadro social
2826 vida social
2831 cultura e religião
2836 protecção social
2841 saúde
2846 urbanismo e construção civil

32 EDUCAÇÃO E COMUNICAÇÃO

3206 educação
3211 ensino
3216 organização do ensino
3221 documentação
3226 comunicação
3231 informação e tratamento da informação
3236 informática e processamento de dados

36 CIÊNCIAS

3606 ciências naturais e aplicadas
3611 ciências humanas

40 EMPRESAS E CONCORRÊNCIA

4006 organização de empresas
4011 tipos de empresa
4016 forma jurídica de sociedade
4021 gestão administrativa
4026 gestão contabilística
4031 concorrência

44 EMPREGO E TRABALHO

4406 emprego
4411 mercado do trabalho
4416 condições e organização do trabalho
4421 administração e remuneração do pessoal
4426 relações laborais e direito do trabalho

48 TRANSPORTES

4806 política de transportes
4811 organização dos transportes
4816 transporte terrestre
4821 transporte marítimo e fluvial
4826 transporte aéreo e espacial

52 MEIO AMBIENTE

5206 política ambiental
5211 meio natural
5216 degradação do ambiente

56 AGRICULTURA, SILVICULTURA E PESCA

5606 política agrícola
5611 produção e estruturas agrícolas
5616 sistema de exploração agrícola
5621 exploração agrícola
5626 meios de produção agrícola
5631 actividade agrícola
5636 floresta
5641 pesca

60 AGRO-ALIMENTAR

6006 produto vegetal
6011 produto animal
6016 produto agrícola transformado
6021 bebidas e açúcar
6026 produto alimentar
6031 agro-alimentar
6036 tecnologia alimentar

64 PRODUÇÃO, TECNOLOGIA E INVESTIGAÇÃO

6406 produção
6411 tecnologia e regulamentação técnica
6416 investigação e propriedade intelectual

66 ENERGIA

6606 política energética
6611 indústrias carbonífera e mineira
6616 indústria petrolífera
6621 indústrias nuclear e eléctrica
6626 energia não poluente

68 INDÚSTRIA

6806 política e estruturas industriais
6811 química
6816 metalurgia e siderurgia
6821 indústria mecânica
6826 electrónica e electrotécnica
6831 construção civil
6836 indústria da madeira
6841 indústria do couro e têxtil
6846 indústrias diversas

72 GEOGRAFIA

7206 Europa
7211 regiões dos Estados-Membros da União Europeia
7216 América
7221 África
7226 Ásia-Oceânia
7231 geografia económica
7236 geografia política
7241 países e territórios ultramarinos

76 ORGANIZAÇÕES INTERNACIONAIS

7606 Nações Unidas
7611 organizações europeias
7616 organizações extra-europeias
7621 organizações mundiais
7626 organizações não governamentais

B

Mapeamento entre EUROVOC e conceitos

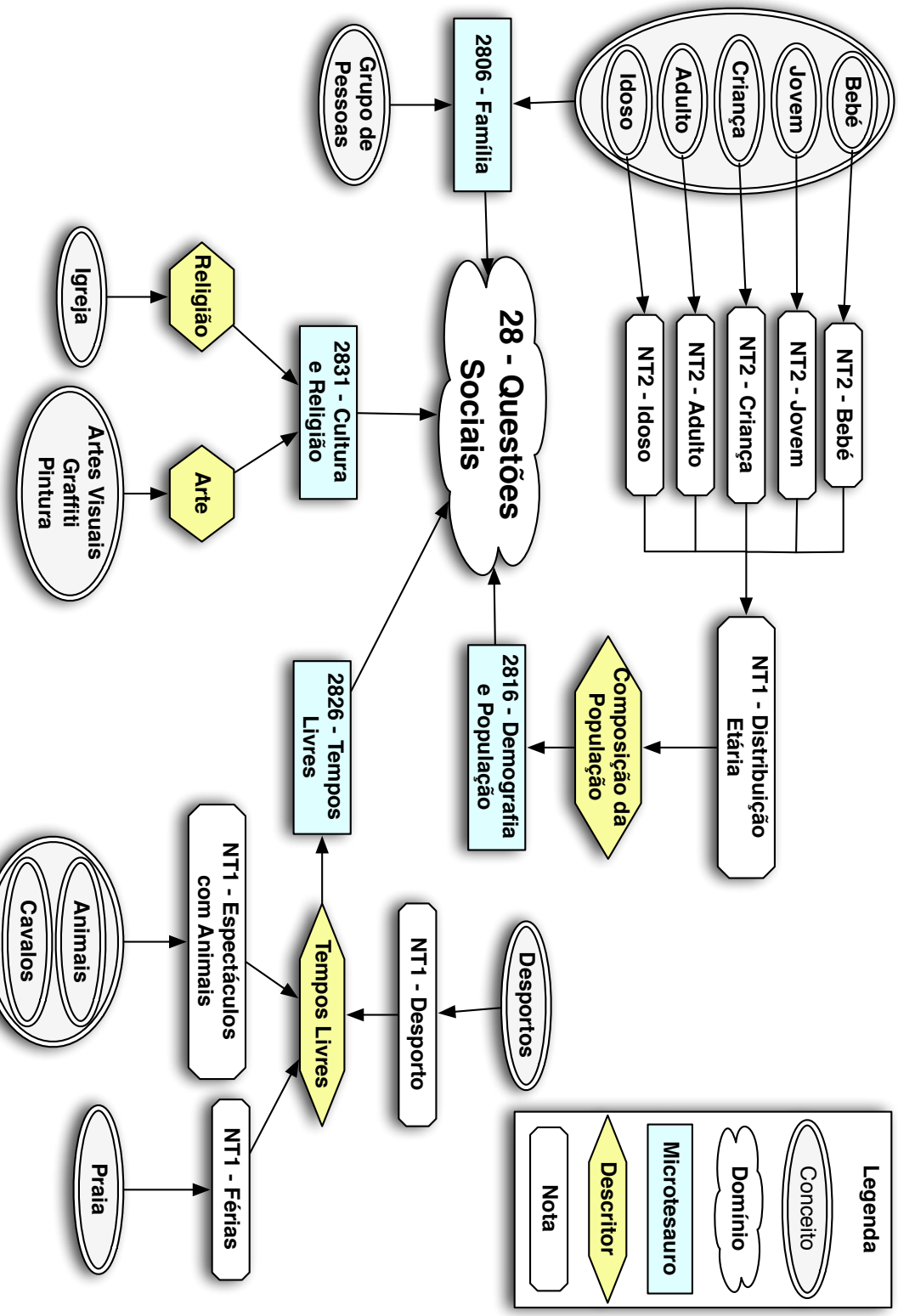


Figura B.1: Questões Sociais

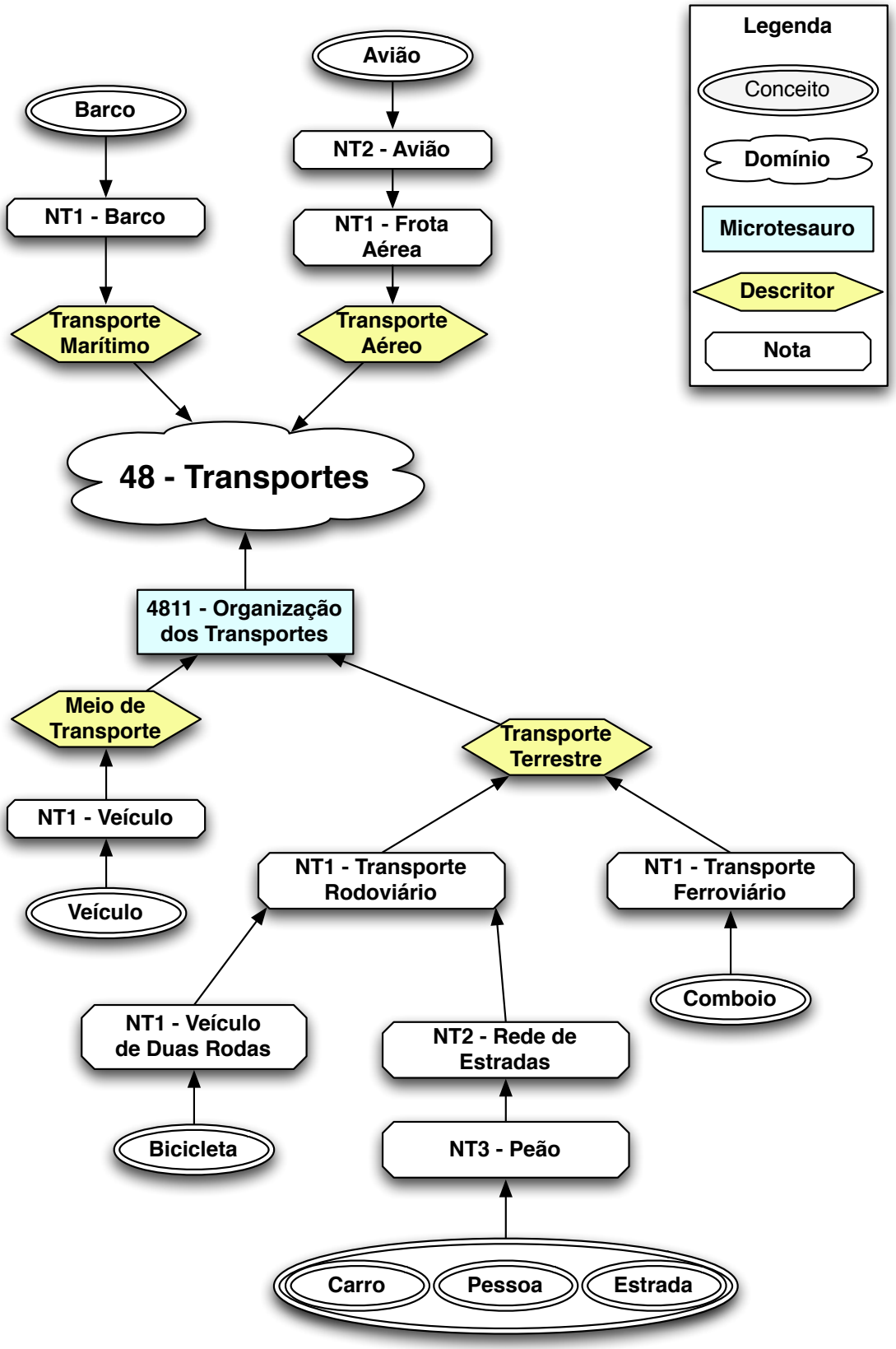


Figura B.2: Transportes

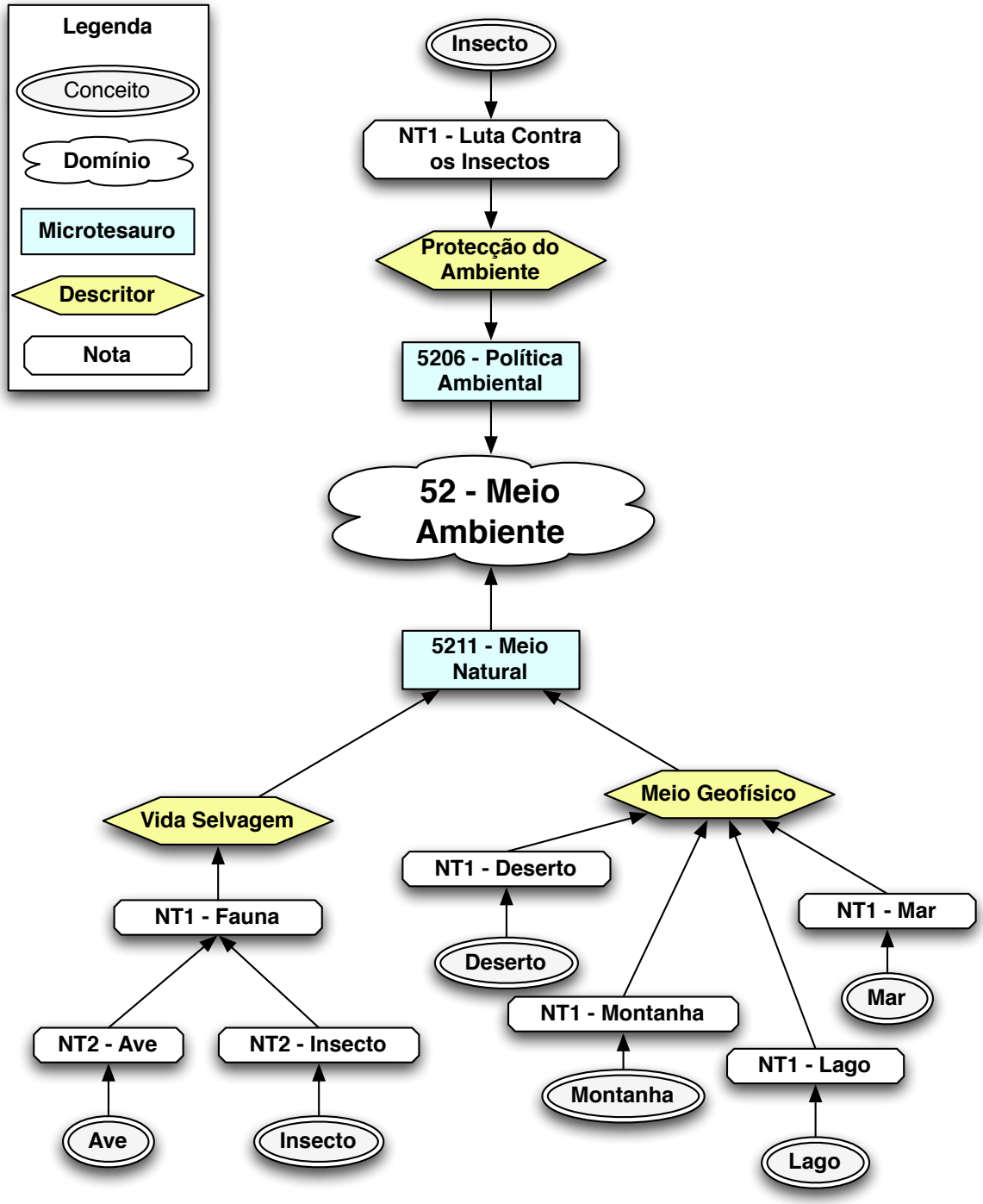


Figura B.3: Meio Ambiente

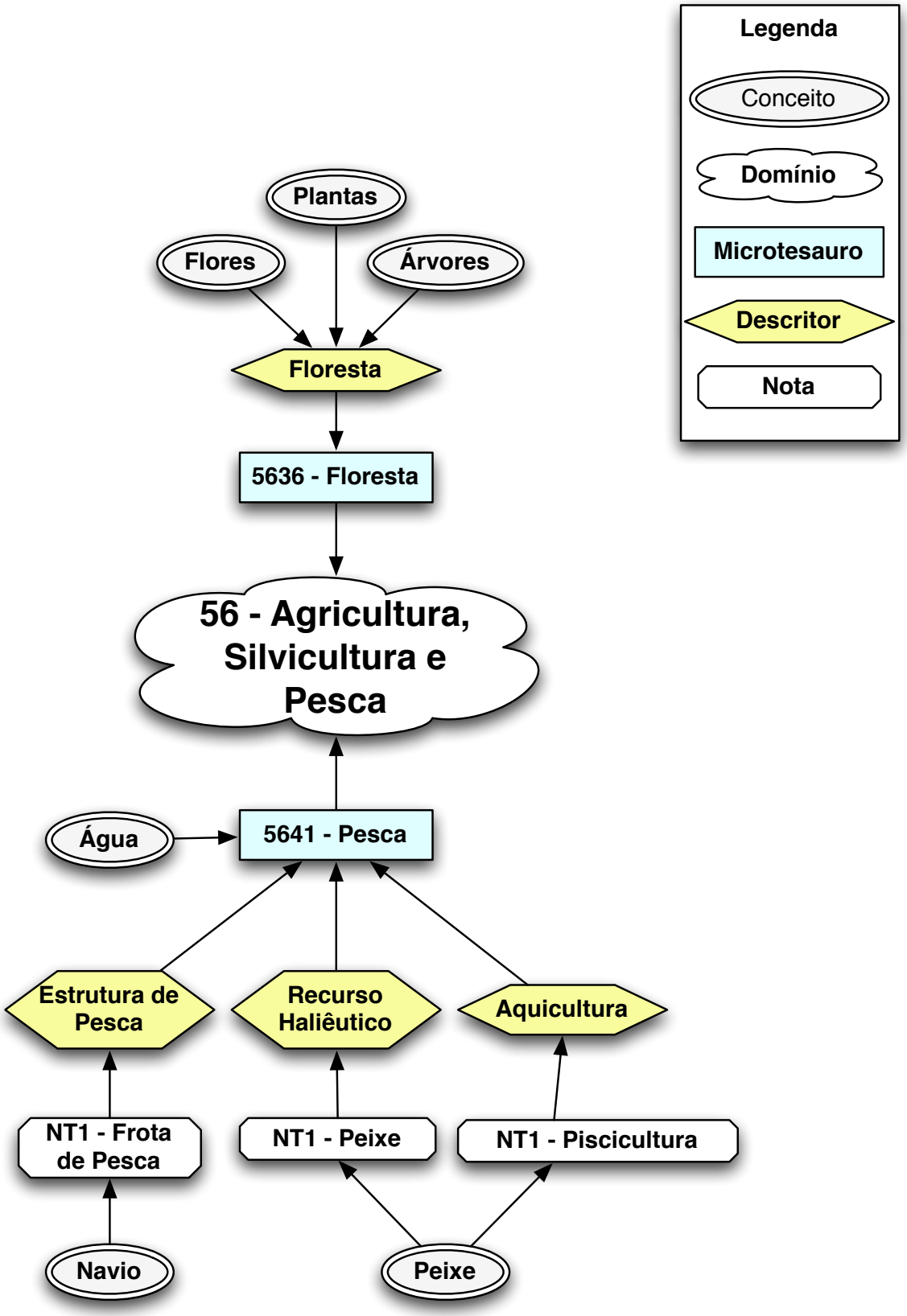


Figura B.4: Agricultura, Silvicultura e Pesca

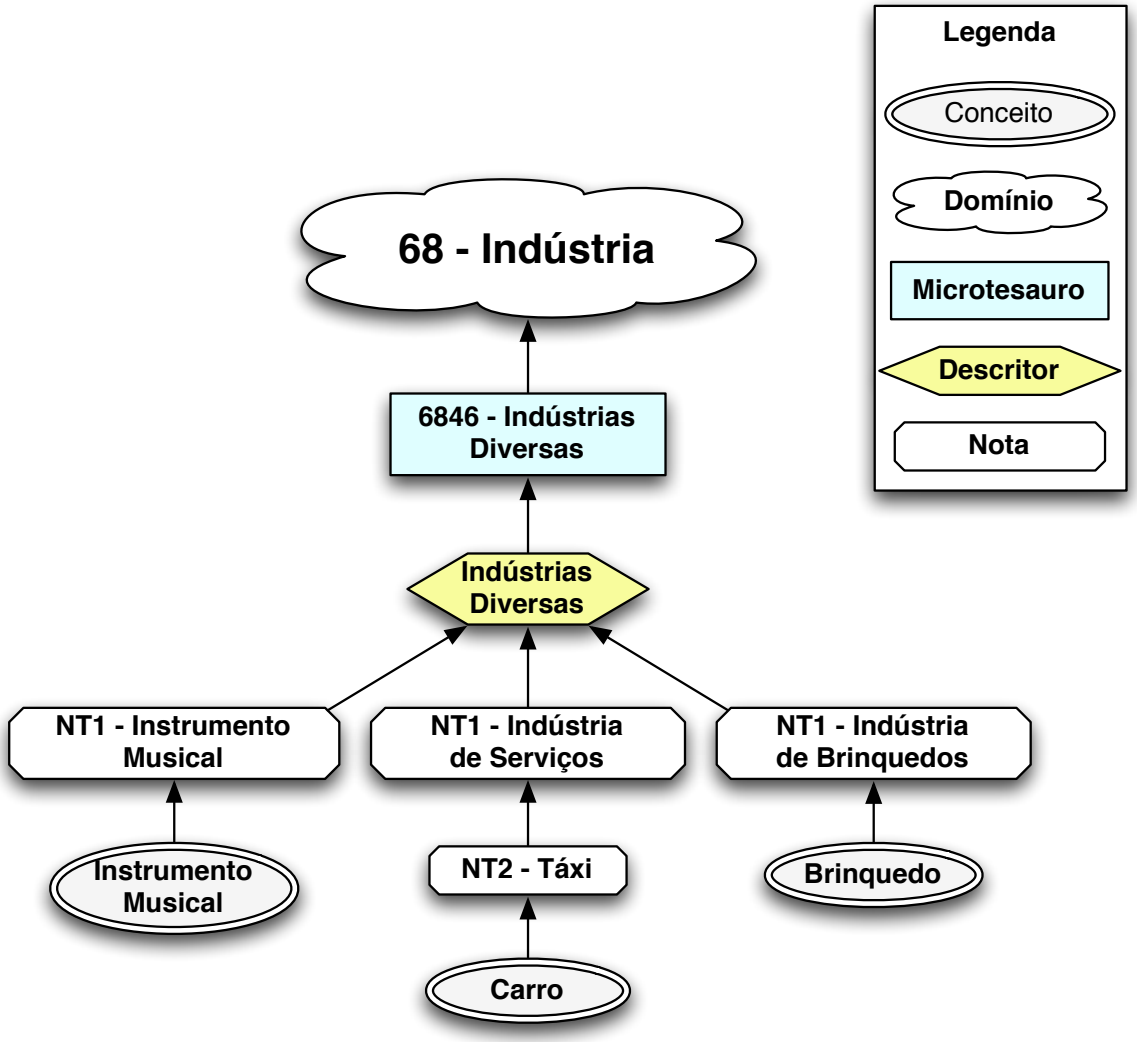


Figura B.5: Indústria

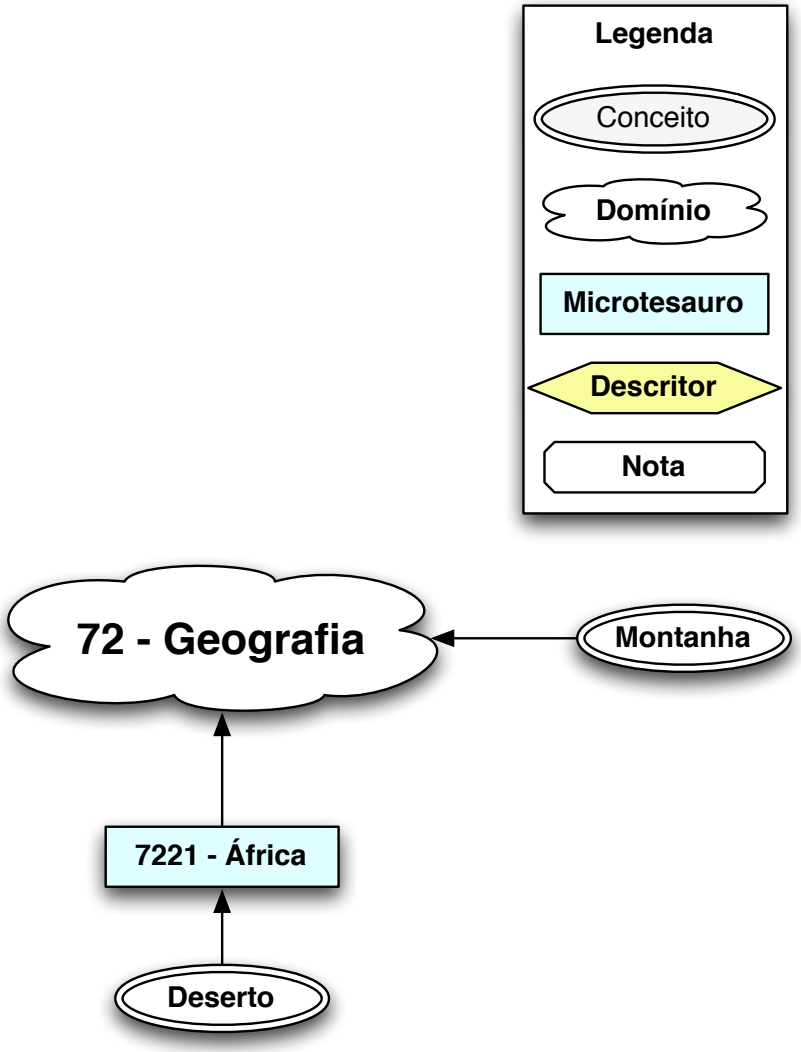


Figura B.6: Geografia

C

Inquérito sobre ViewProcFlow

A aplicação ViewProcFlow pretende auxiliar uma empresa de conteúdos de vídeo, Duvideo, a retirar maior partido do seu arquivo audiovisual. São fornecidos processos para a extracção de metadados. Este questionário foca-se num componente do sistema, que permite ao utilizador visualizar, modificar e avaliar esses metadados.

Tarefas

Por favor, efectue as seguintes tarefas e descreva-nos os resultados obtidos.

Exploração do Ambiente de Visualização

Tarefa 1: Visualize o conteúdo do video "people.mov". Observe as Timelines relativas às faces e às cenas.
Que dificuldades encontrou para cumprir esta tarefa?

Tarefa 2: Escolha o vídeo "little_girl.mov", verifique os metadados extraídos. Peça um novo processamento, modificando os parâmetros disponibilizados.
Que dificuldades encontrou para cumprir esta tarefa?

Tarefa 3: Escolha um vídeo, modifique os dados do video. Adicione uma categoria, escreva um comentário.
Que dificuldades encontrou para cumprir esta tarefa?

Exploração de Ambiente de Pesquisa

Tarefa 4: Efectue uma pesquisa, procure por todos os vídeos que se encontram por validar.
Que dificuldades encontrou para cumprir esta tarefa?

Tarefa 5: Efectue uma pesquisa, inclua uma imagem da biblioteca de imagens como parâmetro de pesquisa.
Que dificuldades encontrou para cumprir esta tarefa?

Exploração de Ambiente de Conceitos e Ontologias

Tarefa 6: Adicione o termo "bus" à categoria "Land Transport".
Que dificuldades encontrou para cumprir esta tarefa?

Tarefa 7: Descreva um novo conceito "car" para ser treinado. Dê-lhe o nome e com recurso às sugestões apresentadas, construa os conjuntos positivos e negativos com as imagens.
Que dificuldades encontrou para cumprir esta tarefa?

Questões gerais relacionadas com as tarefas anteriores

1. Na sua opinião, a interface de visualização encontra-se adequada?

Inadequada				Adequada
1	2	3	4	5

2. Na sua opinião, a interface de pesquisa encontra-se adequada?

Inadequado				Adequado
1	2	3	4	5

3. Na sua opinião, como avalia o desempenho da aplicação em pesquisas sem de imagem?

Lento				Rápido
1	2	3	4	5

4. Na sua opinião, como avalia o desempenho da aplicação em pesquisas com imagem?

Lento				Rápido
1	2	3	4	5

5. Na sua opinião, como avalia os resultados das pesquisas com imagem?

Mau				Bom
1	2	3	4	5

6. Na sua opinião, como avalia a utilização da ontologia/EUROVOC?

Complicada de usar				Simples de usar
1	2	3	4	5

7. Na sua opinião, como avalia a criação de um novo conceito?

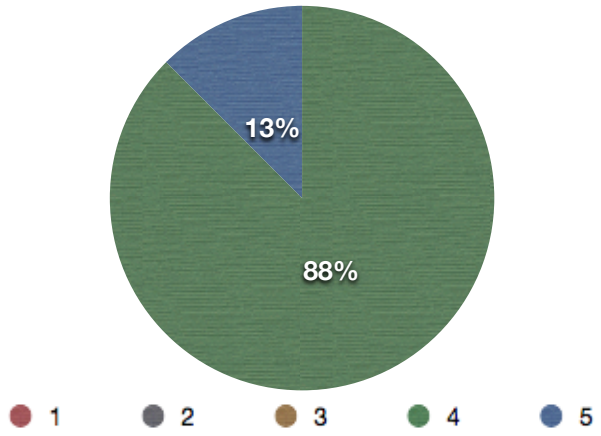
Complicada de usar				Simples de usar
1	2	3	4	5

Comentários e Sugestões

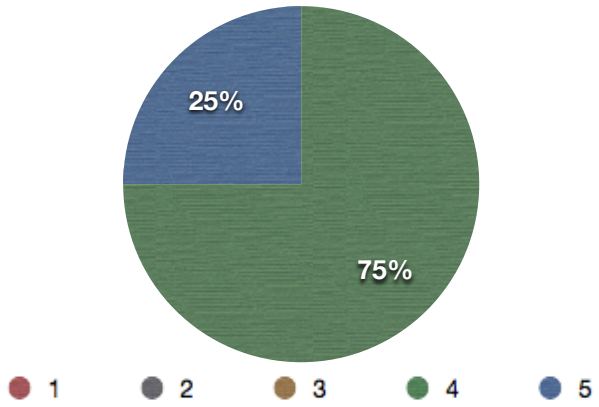
D

Resultados sobre o inquérito à utilização do protótipo

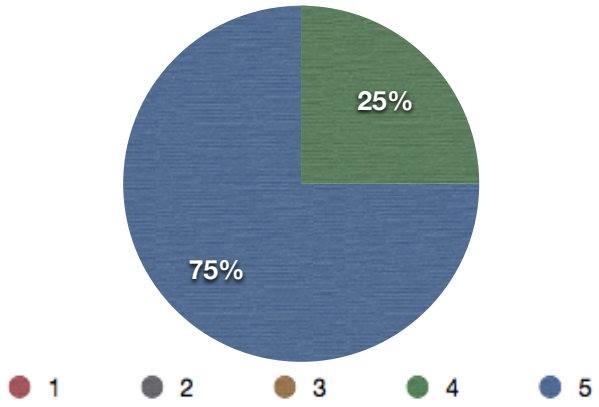
Questão 1) - Na sua opinião, a interface de visualização encontra-se adequada?



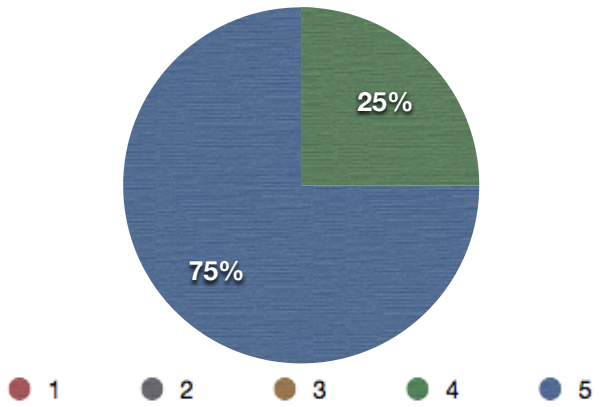
Questão 2) - Na sua opinião, a interface de pesquisa encontra-se adequada?



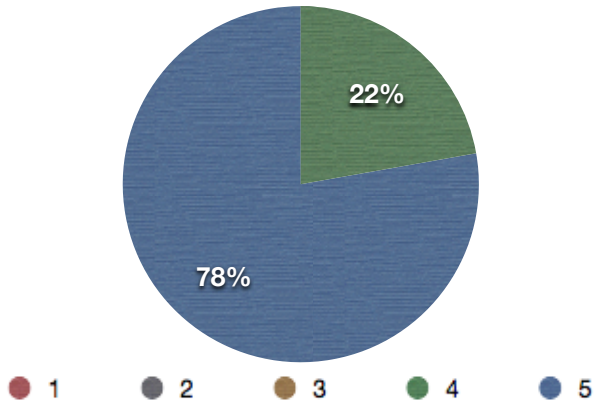
Questão 3) - Na sua opinião, como avalia o desempenho da aplicação em pesquisas sem de imagem?



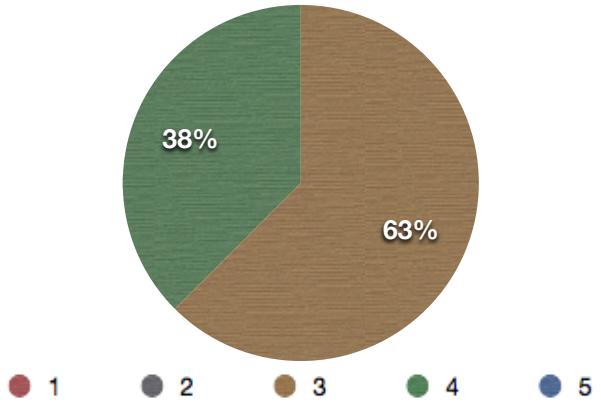
Questão 4) - Na sua opinião, como avalia o desempenho da aplicação em pesquisas com imagem?



Questão 5) - Na sua opinião, como avalia os resultados das pesquisas com imagem?



Questão 6) - Na sua opinião, como avalia a utilização da ontologia/EUROVOC?



Questão 7) - Na sua opinião, como avalia a criação de um novo conceito?